Regression type models for extremal dependence

Regression type models for extremal dependence

L. Mhalla[1] | **M. de Carvalho**[2] | **V. Chavez-Demoulin**[3]

[1]Geneva School of Economics and Management, Université de Genève, 1205 Genève, Switzerland

[2]School of Mathematics, University of Edinburgh, Edinburgh, UK

[3]HEC Lausanne, Université de Lausanne, Lausanne, Switzerland

We propose a vector generalized additive modelling framework for taking into account the effect of covariates on angular density functions in a multivariate extreme value context. The proposed methods are tailored for settings where the dependence between extreme values may change according to covariates. We devise a maximum penalized log-likelihood estimator, discuss details of the estimation procedure, and derive its consistency and asymptotic normality. The simulation study suggests that the proposed methods perform well in a wealth of simulation scenarios by accurately recovering the true covariate-adjusted angular density. Our empirical analysis reveals relevant dynamics of the dependence between extreme air temperatures in two alpine resorts during the winter season.

KEYWORDS

angular density; covariate-adjustment; penalized log-likelihood; statistics of multivariate extremes; vgam.

1 | INTRODUCTION

In this paper, we address an extension of the standard approach for modelling non-stationary univariate extremes to the multivariate setting. In the univariate context, the limiting distribution for the maximum of a sequence of independent and identically distributed random variables, derived by Fisher and Tippett [1928], is given by a generalized extreme value distribution characterized by three parameters: μ (location), σ (scale), and ξ (shape). To take into account the effect of a vector of covariates $\mathbf{x}$, one can let these parameters depend on $\mathbf{x}$, and the resulting generalized extreme value distribution takes the form

$$G_{(\mu_{\mathbf{x}},\sigma_{\mathbf{x}},\xi_{\mathbf{x}})}(z) = \exp\left[-\left\{1+\xi_{\mathbf{x}}\left(\frac{z-\mu_{\mathbf{x}}}{\sigma_{\mathbf{x}}}\right)\right\}_{+}^{-1/\xi_{\mathbf{x}}}\right], \tag{1}$$

where $(a)_+ = \max\{0, a\}$; see Coles [2001, ch. 6], Pauli and Coles [2001], Chavez-Demoulin and Davison [2005], Yee and Stephenson [2007], Wang and Tsai [2009], Eastoe and Tawn [2009], and Chavez-Demoulin and Davison [2005] for related approaches.

In the multivariate context, consider $\mathbf{Y}^i = \left(Y_1^i, \ldots, Y_d^i\right)^\top$ independent and identically distributed random vectors with joint distribution F, and unit Fréchet marginal distribution functions $F_j(y) = \exp(-1/y)$, for $y > 0$. Pickands' representation theorem [Coles, 2001, Theorem 8.1] states that the law of the standardized componentwise maxima, $\mathbf{M}_n = n^{-1} \max\{\mathbf{Y}^1, \ldots, \mathbf{Y}^n\}$, converges in distribution to a multivariate extreme value distribution, $G_H(\mathbf{z}) = \exp\{-V_H(\mathbf{z})\}$, with

$$V_H(\mathbf{z}) = \int_{S_d} \max\left(\frac{w_1}{z_1}, \ldots, \frac{w_d}{z_d}\right) \mathrm{d}H(\mathbf{w}), \quad \mathbf{z} \in (0, \infty)^d. \tag{2}$$

Here H is the so-called angular measure, that is, a positive finite measure on the unit simplex $S_d = \left\{(w_1, \ldots, w_d) \in [0,1]^d : w_1 + \cdots + w_d = 1\right\}$ that needs to obey

$$\int_{S_d} w_j \, \mathrm{d}H(\mathbf{w}) = 1, \quad j = 1, \ldots, d. \tag{3}$$

The function $V(\mathbf{z}) \equiv V_H(\mathbf{z})$, is the so-called exponent measure and is continuous, convex, and homogeneous of order -1, i.e., $V(t\mathbf{z}) = t^{-1}V(\mathbf{z})$ for all $t > 0$.

The class of limiting distributions of multivariate extreme values yields an infinite number of possible parametric representations [Coles, 2001, ch. 8], as the validity of a multivariate extreme value distribution is conditional on its angular measure H satisfying the moment constraint (3). Therefore, most literature has focused on the estimation of the extremal dependence structures described by spectral measures or equivalently angular densities [Boldi and Davison, 2007, Einmahl et al., 2009, de Carvalho et al., 2013, Sabourin and Naveau, 2014, Hanson et al., 2017]. Related quantities, such as the Pickands dependence function [Pickands, 1981] and the stable tail dependence function [Huang, 1992, Drees and Kaufmann, 1998], were investigated by many authors [Einmahl et al., 2006, Gudendorf and Segers, 2012, Wadsworth and Tawn, 2013, Marcon et al., 2016]. A wide variety of parametric models for the spectral density that allow flexible dependence structures were proposed [Kotz and Nadarajah, 2000, sec. 3.4].

However, few papers were able to satisfactorily address the challenging but incredibly relevant setting of modelling nonstationarity at joint extreme levels. Some exceptions include de Carvalho and Davison [2014], who proposed a nonparametric approach, where a family of spectral densities is constructed using exponential tilting. Castro and de Carvalho [2017] developed an extension of this approach based on covariate-varying spectral densities. However, these approaches are limited to replicated one-way ANOVA types of settings. de Carvalho [2016] advocated the use of covariate-adjusted angular densities, and Escobar-Bach et al. [2016] discussed estimation—in the bivariate and covariate-dependent framework—of the Pickands dependence function based on local estimation with a minimum density power divergence criterion. Recently, Mhalla et al. [2017] constructed in a nonparametric framework smooth models for predictor-dependent Pickands dependence functions based on generalized additive models, whereas Castro et al. [2018] proposed nonparametric regression methods for predictor-dependent angular measures; a key advantage of our method is that it can be used for modelling a high number (with limitation given by the sample size) of covariates of any type (from categorical to continuous), and it combines the flexibility of GAM along with a parametric specification to effectively learn about the dynamics governing the extremal dependence structure.

Our approach is based on a non-linear model for covariate-varying extremal dependences. Specifically, we develop a vector generalized additive model that flexibly allows the extremal dependence to change with a set of covariates, but—keeping in mind that extreme values are scarce—it borrows strength from a parametric assumption. In other words, the goal is to develop a regression model for the extremal dependence through the parametric specification of an extremal dependence structure and then to model the parameters of that structure through a vector generalized

additive model (VGAM) [Yee and Wild, 1996, Yee, 2015]. One major advantage over existing methods is that our model may be used for handling an arbitrary number of dimensions and covariates of different types, and it is straightforward to implement, as illustrated in the R code [R Development Core Team, 2016] that can be downloaded from `https://github.com/lindamhalla/Regression_type_models_for_extremal_dependence`.

The remainder of this paper is organized as follows. In Section 2 we introduce the proposed model for covariate-adjusted extremal dependences. In Section 3 we develop our penalized likelihood approach and give details on the asymptotic properties of our estimator. In Section 4 we assess the performance of the proposed methods. An application to extreme temperatures in the Swiss Alps is given in Section 5. We close the paper in Section 6 with a discussion.

2 | FLEXIBLE COVARIATE-ADJUSTED ANGULAR DENSITIES

Section 2.1 offers preparations and background, Section 2.2 introduces the proposed vector generalized additive model; parametric examples and further comments are included in Section 2.3.

2.1 Statistics of multivariate extremes: preparations and background

The functions H and V in (2) can be used to describe the structure of dependence between the extremes, as in the case of independence between the extremes, where $V(\mathbf{z}) = \sum_{j=1}^{d} 1/z_j$, and in the case of perfect extremal dependence, where $V(\mathbf{z}) = \max\{1/z_1, \ldots, 1/z_d\}$. As a consequence, if H is differentiable with angular density denoted h, the more mass around the barycenter of S_d, $(d^{-1}, \ldots, d^{-1})$, the higher the level of extremal dependence. Further insight into these measures may be obtained by considering the point process $P_n = \{n^{-1}\mathbf{Y}^i : i = 1, \ldots, n\}$. Following de Haan and Resnick [1977] and Resnick [1987, Section 5.3], as $n \to \infty$, P_n converges to a non-homogeneous Poisson point process P defined on $[0, \infty) \setminus \{0\}$ with a mean measure μ that verifies

$$\mu(A_{\mathbf{y}}) = V(\mathbf{y}),$$

where $A_{\mathbf{y}} = \mathbb{R}^d \setminus ([-\infty, y_1] \times \cdots \times [-\infty, y_d])$.

There are two representations of the intensity measure of the limiting Poisson point process P that will be handy for our purposes. First, it holds that

$$\mu(\mathrm{d}\mathbf{y}) = -V_{1:d}(\mathbf{y})\, \mathrm{d}\mathbf{y}, \tag{4}$$

with $V_{1:d}$ being the derivative of V with respect to all its arguments [Resnick, 1987, Section 5.4]. Second, another useful factorization of the intensity measure $\mu(\mathrm{d}\mathbf{y})$, called the spectral decomposition, can be obtained using the following decomposition of the random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)^{\mathsf{T}}$ into radial and angular coordinates,

$$(R, \mathbf{W}) = \left(\|\mathbf{Y}\|, \frac{\mathbf{Y}}{\|\mathbf{Y}\|} \right), \tag{5}$$

where $\|\cdot\|$ denotes the L_1-norm. It can be shown that [Beirlant et al., 2004, Section 8.2.3] the limiting intensity

measure factorizes across radial and angular components as follows:

$$\mu(\mathrm{d}\mathbf{y}) = \mu(\mathrm{d}r \times \mathrm{d}\mathbf{w}) = \frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w}).$$

Here r measures the distance from the origin and each component of $\mathbf{w}$ measures angles on a $[0, 1]$ scale. The spectral decomposition (5) allows the separation of the marginal and the dependence parts in the multivariate extreme value distribution G_H, with the margins being unit Fréchet and the dependence structure being described by the angular measure H.

The inference approach that we build on in this work was developed by Coles and Tawn [1991] and is based on threshold excesses; see Huser et al. [2016] for a detailed review of likelihood estimators for multivariate extremes. The set of extreme events is defined as the set of observations with radial components exceeding a high fixed threshold, that is, the observations belonging to the extreme set,

$$E_{\mathbf{r}} = \left\{(y_1, \ldots, y_d) \in (0, \infty)^d : \sum_{j=1}^{d} \frac{y_j}{r_j} > 1\right\},$$

with $\mathbf{r} = (r_1, \ldots, r_d)$ being a large threshold vector. Since the points $n^{-1}\mathbf{Y}^i$ are mapped to the origin for non-extreme observations, the threshold $\mathbf{r}$ needs to be sufficiently large for the Poisson approximation to hold. Note that, $\mathbf{Y}^i \in E_{\mathbf{r}}$, if and only if,

$$R_i = \|\mathbf{Y}^i\| > \left(\sum_{j=1}^{d} \frac{\omega_{i,j}}{r_j}\right)^{-1}, \quad \text{where } \omega_{i,j} = \frac{Y_j^i}{R_i}.$$

Hence, the expected number of points of the Poisson process P located in the extreme region $E_{\mathbf{r}}$ is

$$\begin{aligned}
\mu(E_{\mathbf{r}}) &= \int_{S_d} \int_{\left(\sum_{j=1}^{d} \frac{w_j}{r_j}\right)^{-1}}^{\infty} \frac{\mathrm{d}r}{r^2}\,\mathrm{d}H(\mathbf{w}) \\
&= \int_{S_d} \left(\sum_{j=1}^{d} \frac{w_j}{r_j}\right) \mathrm{d}H(\mathbf{w}) \\
&= \sum_{j=1}^{d} \frac{1}{r_j} \int_{S_d} w_j\,\mathrm{d}H(\mathbf{w}) = \sum_{j=1}^{d} \frac{1}{r_j}. \qquad (6)
\end{aligned}$$

Now, we can explicitly formulate the Poisson log-likelihood over the set E_r,

$$\ell_{E_{\mathbf{r}}}(\boldsymbol{\theta}) = -\mu(E_{\mathbf{r}}) + \sum_{i=1}^{n_{\mathbf{r}}} \log\left\{\mu(\mathrm{d}R_i \times \mathrm{d}\mathbf{w}_i)\right\}, \qquad (7)$$

where $\mathbf{w}_i = (\omega_{i,1}, \ldots, \omega_{i,d})$, $\boldsymbol{\theta}$ represents the p-vector of parameters of the measure μ and $n_{\mathbf{r}}$ represents the number of reindexed observations in the extreme set $E_{\mathbf{r}}$. Using (6), the first term in (7) can be omitted when maximizing the Poisson log-likelihood, which, using (4), boils down to

$$\ell_{E_{\mathbf{r}}}(\boldsymbol{\theta}) \equiv \sum_{i=1}^{n_{\mathbf{r}}} \log\left\{-V_{1:d}(\mathbf{Y}^i; \boldsymbol{\theta})\right\}. \qquad (8)$$

Thanks to the differentiability of the exponent measure V and the support of the angular measure H in the unit simplex S_d, we can use the result of Coles and Tawn [1991, Theorem 1] that relates the angular density to the exponent measure via

$$V_{1:d}(\mathbf{y};\boldsymbol{\theta}) = -\|\mathbf{y}\|^{-(d+1)} h\left(\frac{y_1}{\|\mathbf{y}\|},\ldots,\frac{y_d}{\|\mathbf{y}\|};\boldsymbol{\theta}\right)$$

and reformulate the log-likelihood (8) as follows

$$\ell_{E_{\mathbf{r}}}(\boldsymbol{\theta}) \quad \equiv \quad -(d+1)\sum_{i=1}^{n_{\mathbf{r}}} \log\|\mathbf{Y}^i\| + \sum_{i=1}^{n_{\mathbf{r}}} \log\left\{h\left(\frac{Y_1^i}{\|\mathbf{Y}^i\|},\ldots,\frac{Y_d^i}{\|\mathbf{Y}^i\|};\boldsymbol{\theta}\right)\right\}. \tag{9}$$

2.2 Vector generalized additive models for covariate-adjusted angular densities

Our starting point for modelling is an extension of (1) to the multivariate setting. Whereas the model in (1) is based on indexing the parameters of the univariate extreme value distribution with a regressor, here we index the parameter (H) of a multivariate extreme value distribution (G_H) with a regressor $\mathbf{x} = (x_1,\ldots,x_q)^{\mathsf{T}} \in \mathcal{X} \subset \mathbb{R}^q$. Our target object of interest is thus given by a family of covariate-adjusted angular measures $H_{\mathbf{x}}$ obeying

$$\int_{S_d} w_j \mathrm{d}H_{\mathbf{x}}(\mathbf{w}) = 1, \quad j = 1,\ldots,d.$$

Of particular interest is the setting where $H_{\mathbf{x}}$ is differentiable, in which case the covariate-adjusted angular density can be defined as $h_{\mathbf{x}}(\mathbf{w}) = \mathrm{d}H_{\mathbf{x}}/\mathrm{d}\mathbf{w}$. This yields a corresponding family of covariate-indexed multivariate extreme value distributions

$$G_{\mathbf{x}}(\mathbf{z}) = \exp\left\{-\int_{S_d} \max\left(\frac{w_1}{z_1},\ldots,\frac{w_d}{z_d}\right)\mathrm{d}H_{\mathbf{x}}(\mathbf{w})\right\}.$$

Other natural objects depending on $G_{\mathbf{x}}$ can be readily defined, such as the covariate-adjusted extremal coefficient, $\vartheta_{\mathbf{x}}$, which solves

$$G_{\mathbf{x}}(z\mathbf{1}_d) = \exp(-\vartheta_{\mathbf{x}}/z), \quad z > 0, \tag{10}$$

where $\mathbf{1}_d$ is a d-vector of ones. Here, $\vartheta_{\mathbf{x}}$ ranges from 1 to d, and the closer $\vartheta_{\mathbf{x}}$ is to one, the closer we get to the case of complete dependence at that value of the covariate.

The question of interest is thus: How can we learn about $h_{\mathbf{x}}$ from the data? Suppose we observe the regression data $\{(\mathbf{x}^i,\mathbf{Y}^i)\}_{i=1}^n$, with $(\mathbf{x}^i,\mathbf{Y}^i) \in \mathcal{X}\times\mathbb{R}^d$, and where we assume that $\mathbf{Y}^i = \left(Y_1^i,\ldots,Y_d^i\right)^{\mathsf{T}}$ are conditionally independent random vectors with unit Fréchet marginal distributions and a joint distribution in the maximum domain of attraction of a multivariate extreme value distribution with angular density $h_{\mathbf{x}_i}$. Using a similar approach as in Section 2.1, we convert the raw sample into a pseudo-sample of cardinality $n_{\mathbf{r}}$,

$$\{(\mathbf{x}^i,\mathbf{Y}^i) : \mathbf{Y}^i \in E_{\mathbf{r}}\},$$

and use the latter reindexed data to learn about $h_{\mathbf{x}}$ through the set of angular observations $\{\mathbf{w}_i\}_{i=1}^{n_{\mathbf{r}}}$, where $\mathbf{w}_i =$

$\mathbf{Y}^i / \|\mathbf{Y}^i\|$ for $\mathbf{Y}^i \in E_{\mathbf{r}}$.

Without loss of generality, we restrain ourselves to the bivariate extreme value framework ($d = 2$), so that

$$h_{\mathbf{x}}\left(\frac{Y_1^i}{\|\mathbf{Y}^i\|}, \frac{Y_2^i}{\|\mathbf{Y}^i\|}\right) = h_{\mathbf{x}}(w_i, 1 - w_i) \equiv h_{\mathbf{x}}(w_i), \quad \text{for } w_i \in [0,1], \quad i = 1, \ldots, n_{\mathbf{r}},$$

that is, the dimension of the angular observations w_i is $M = d - 1 = 1$. We model $h_{\mathbf{x}}(\cdot)$ using $h(\cdot; \boldsymbol{\theta}_{\mathbf{x}})$, where the parameter underlying the dependence structure

$$\begin{aligned} \boldsymbol{\theta}_{\mathbf{x}} &= (\theta_{1\mathbf{x}^1}, \ldots, \theta_{1\mathbf{x}^{n_{\mathbf{r}}}}, \ldots, \theta_{p\mathbf{x}^1}, \ldots \theta_{p\mathbf{x}^{n_{\mathbf{r}}}})^\top \in \mathbb{R}^{pn_{\mathbf{r}}}, \\ \mathbf{x} &= (\mathbf{x}^1, \ldots, \mathbf{x}^{n_{\mathbf{r}}})^\top \in \mathcal{X}^{n_{\mathbf{r}}} = \left(\mathcal{X}_1 \times \cdots \times \mathcal{X}_q\right)^{n_{\mathbf{r}}} \subseteq \mathbb{R}^{qn_{\mathbf{r}}}, \end{aligned}$$

is specified through a vector generalized additive model (VGAM) [Yee and Wild, 1996]. Specifically, we model $h_{\mathbf{x}}(w)$ using a fixed family of parametric extremal dependence structures $h(w; \boldsymbol{\theta}_{\mathbf{x}})$ with a covariate-dependent set of parameters $\boldsymbol{\theta}_{\mathbf{x}}$. To learn about $\boldsymbol{\theta}_{\mathbf{x}}$ from the pseudo-sample, we use a vector generalized additive model, which takes the form

$$\boldsymbol{\eta}(\mathbf{x}) \equiv \boldsymbol{\eta} = \mathbf{H}_0 \boldsymbol{\beta}_{[0]} + \sum_{k=1}^{q} \mathbf{H}_k \mathbf{f}_k(\mathbf{x}_k). \tag{11}$$

Here,

- $\boldsymbol{\eta} = \mathbf{g}(\boldsymbol{\theta}_{\mathbf{x}}) = \left(g_1(\theta_{1x^1}), \ldots, g_1(\theta_{1x^{n_{\mathbf{r}}}}), \ldots, g_p(\theta_{px^1}), \ldots, g_p(\theta_{px^{n_{\mathbf{r}}}})\right)^\top$ is the vector of predictors and g_l is a link function that ensures that $\theta_{l\cdot}$ is well defined, for $l = 1, \ldots, p$,
- $\boldsymbol{\beta}_{[0]}$ is a $pn_{\mathbf{r}}$-vector of intercepts, with p distinct values each repeated $n_{\mathbf{r}}$ times,
- $\mathbf{x}_k = \left(x_k^1, \ldots, x_k^{n_{\mathbf{r}}}\right)^\top \in \mathcal{X}_k^{n_{\mathbf{r}}}$, for $k = 1, \ldots, q$,
- $\mathbf{f}_k = (\mathbf{f}_{k,1}, \ldots, \mathbf{f}_{k,p})^\top$, where $\mathbf{f}_{k,l} = (f_{k,l}(x_k^1), \ldots, f_{k,l}(x_k^{n_{\mathbf{r}}}))^\top$, and $f_{k,l} : \mathcal{X}_k \to \mathbb{R}$ are smooth functions supported on $\mathcal{X}_k$, for $k = 1, \ldots, q$ and $l = 1, \ldots, p$, and
- $\mathbf{H}_k$ are $pn_{\mathbf{r}} \times pn_{\mathbf{r}}$ constraint matrices, for $k = 0, \ldots, q$.

Some comments on the advantages of the VGAM specification are in order. As can be seen from (11) this specification allows us to handle as many covariates as we are required to, whereas this is not as straightforward in other approaches (e.g., kernel regression). The VGAM specification is also compelling in terms of interpretability as it allows us to untangle how each dependence parameter changes with covariates. Yet another benefit of this paradigm, is that the specification (11) is tailored for handling an arbitrary number of covariate-dependent parameters as well as an arbitrary number of covariates of different types. Finally, the VGAM offers a large flexibility of covariate dependence functional form taking advantage of a rich variety of smoothing models that use different basis smoothers such as the cubic or the cyclic cubic splines and the thin plate splines (see Wood [2017], Chapter 5).

By allowing the parameters of the extremal dependence structure to depend on a set of covariates, the proposed specification (11) flexibly allows the dependence between the extreme values to change as a function of covariates. The constraint matrices $\mathbf{H}_k$ are important quantities in the VGAM (11) that allow the tuning of the effects of the covariates on each of the $pn_{\mathbf{r}}$ components of $\boldsymbol{\eta}$. For example, in Example 2.4 (see Section 2.3), one might want to impose the same smooth effect of a covariate on each of the $\binom{3}{2}$ pairwise dependence parameters and at the same time restrict the effect of this covariate to be zero on the global dependence parameter. To avoid clutter in the notation,

we assume from now on that $\mathbf{H}_k = \mathbf{I}_{pn_\mathbf{r} \times pn_\mathbf{r}}$, for $k = 0, \ldots, q$.
The smooth functions $f_{k,l}$ are written as linear combinations of B-spline basis functions

$$f_{k,l}(x_k^i) = \sum_{s=1}^{d_k} \beta_{[kl]_s} B_{s,\tilde{q}}(x_k^i), \quad k = 1, \ldots, q, \quad l = 1, \ldots, p, \quad i = 1, \ldots, n_\mathbf{r},$$

where $B_{s,\tilde{q}}$ is the sth B-spline of order $\tilde{q}$ and $d_k = \tilde{q} + m_k$, with m_k the number of internal equidistant knots for $\mathbf{x}_k$ [Yee, 2015, Section 2.4.5]. To ease the notational burden, we suppose without loss of generality that $d_k \equiv \tilde{d}$, for $k = 1, \ldots, q$, and define

$$\boldsymbol{\beta}_{[k]} = \left(\beta_{[k1]_1}, \ldots, \beta_{[k1]_{\tilde{d}}}, \ldots, \beta_{[kp]_1}, \ldots, \beta_{[kp]_{\tilde{d}}}\right)^\top \in \mathbb{R}^{\tilde{d}p}.$$

Therefore, the VGAM (11), with identity constraint matrices $\mathbf{H}_k$, can be written as

$$\boldsymbol{\eta} = \boldsymbol{\beta}_{[0]} + \sum_{k=1}^{q} \mathbf{X}_{[k]} \boldsymbol{\beta}_{[k]} = \mathbf{X}_{\text{VAM}} \boldsymbol{\beta}, \tag{12}$$

where

$$\begin{cases} \boldsymbol{\beta} = \left(\boldsymbol{\beta}_{[0]} \quad \boldsymbol{\beta}_{[1]} \quad \cdots \quad \boldsymbol{\beta}_{[q]}\right)^\top \in \mathbf{B} \subset \mathbb{R}^{p(1+q\tilde{d})}, \\ \mathbf{X}_{\text{VAM}} = \left(\mathbf{1}_{pn_\mathbf{r} \times p} \quad \mathbf{X}_{[1]} \quad \cdots \quad \mathbf{X}_{[q]}\right) \in \mathbb{R}^{pn_\mathbf{r} \times \{p(1+q\tilde{d})\}}, \end{cases}$$

for some $pn_\mathbf{r} \times \tilde{d}p$ submatrices $\mathbf{X}_{[k]}$, $k = 1, \ldots, q$. The vector of parameters to be estimated in the VGAM (12) is $\boldsymbol{\beta}$. The specification in (12) makes it possible to simultaneously fit ordinary generalized additive models [Wood, 2017] in each component of the vector of parameters $\boldsymbol{\theta}_\mathbf{x}$, hence avoiding any non orthogonality-related issues that could arise if the p components were to be treated separately [Chavez-Demoulin and Davison, 2005]. Finally, if the dimension M of the response vector of angular observations w_i is greater than one ($d > 2$), then the vector of predictors $\boldsymbol{\eta}$ will instead be a $Mpn_\mathbf{r}$-vector and the dimensions of the related quantities in (12) will change accordingly.

2.3 Examples

Some parametric models [Hüsler and Reiss, 1989, Tawn, 1990, Coles and Tawn, 1991, Cooley et al., 2010] are used below to illustrate the concept of covariate-adjusted angular densities and of covariate-adjusted extremal coefficients, and we focus on the bivariate and trivariate settings for the sake of illustrating ideas. To develop insight and intuition on these models, see Figures 1 and 2.

Example 2.1 (Logistic angular surface) Let

$$h_\mathbf{x}(w) = (1/\alpha_\mathbf{x} - 1)\,\{w(1-w)\}^{-1-1/\alpha_\mathbf{x}}\,\{w^{-1/\alpha_\mathbf{x}} + (1-w)^{-1/\alpha_\mathbf{x}}\}^{\alpha_\mathbf{x}-2}, \quad w \in (0,1),$$

with $\alpha : \mathcal{X} \subset \mathbb{R}^q \to (0,1]$. In Figure 1 (left) we represent the case $\alpha_x = \exp\{\eta(x)\}/[1 + \exp\{\eta(x)\}]$, with $\eta(x) = x^2 - 0.5x - 1$ and $x \in \mathcal{X} = [0.1, 2]$. This setup corresponds to be transitioning between a case of relatively high extremal dependence (lower values of x) to a case where we approach asymptotic independence (higher values of x).

Example 2.2 (Dirichlet angular surface) Let

$$h_{\mathbf{x}}(w) = \frac{\alpha_{\mathbf{x}}\beta_{\mathbf{x}}\Gamma(\alpha_{\mathbf{x}}+\beta_{\mathbf{x}}+1)(\alpha_{\mathbf{x}}w)^{\alpha_{\mathbf{x}}-1}\{\beta_{\mathbf{x}}(1-w)\}^{\beta_{\mathbf{x}}-1}}{\Gamma(\alpha_{\mathbf{x}})\Gamma(\beta_{\mathbf{x}})\{\alpha_{\mathbf{x}}w+\beta_{\mathbf{x}}(1-w)\}^{\alpha_{\mathbf{x}}+\beta_{\mathbf{x}}+1}}, \quad w \in (0,1),$$

with $\alpha : \mathcal{X} \subset \mathbb{R}^q \to (0,\infty)$ and $\beta : \mathcal{X} \subset \mathbb{R}^q \to (0,\infty)$. In Figure 1 (middle) we consider the case $\alpha_x = \exp(x)$ and $\beta_x = x^2$, with $x \in [0.9, 3]$. Note the different schemes of extremal dependence induced by the different values of the covariate x as well as the asymmetry of the angular surface underlying this model.

Example 2.3 (Hüsler–Reiss angular surface) Let

$$h_{\mathbf{x}}(w) = \frac{\lambda_{\mathbf{x}}}{w(1-w)^2(2\pi)^{1/2}} \exp\left(-\frac{\left[2+\lambda_{\mathbf{x}}^2\log\{w/(1-w)\}\right]^2}{8\lambda_{\mathbf{x}}^2}\right) \quad w \in (0,1),$$

where $\lambda : \mathcal{X} \subset \mathbb{R}^q \to (0,\infty)$. In Figure 1 (right) we consider the case $\lambda_x = \exp(x)$, with $x \in [0.1, 2]$. Under this specification, lower values of x correspond to lower levels of extremal dependence, whereas higher values of x correspond to higher levels of extremal dependence.

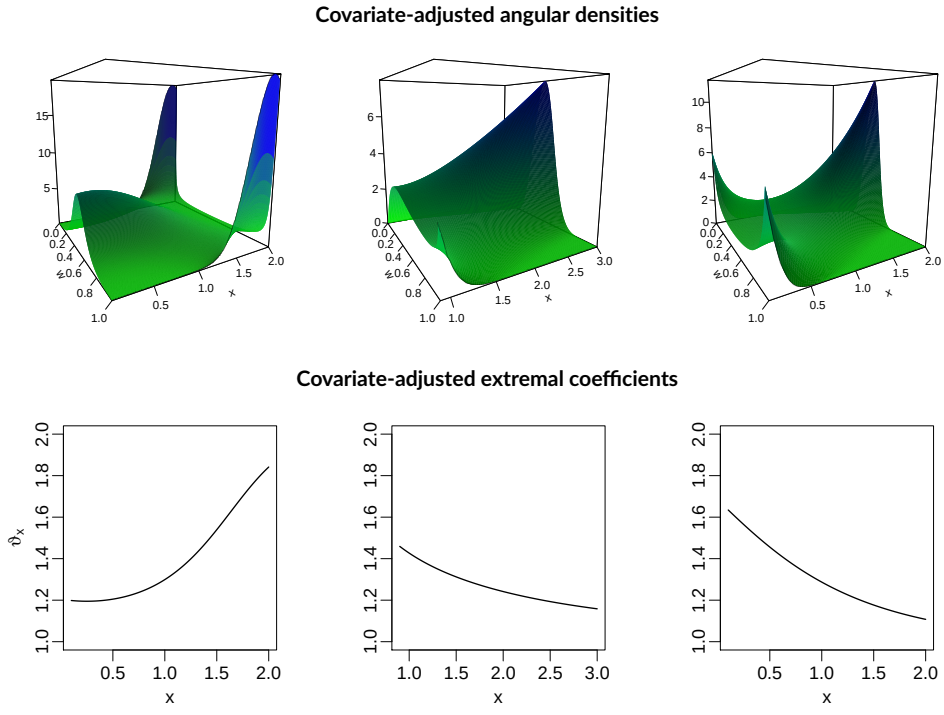

FIGURE 1 Covariate-adjusted angular densities and extremal coefficients of logistic (left panels), Dirichlet (middle panels), and Hüsler–Reiss (right panels) models, corresponding, respectively, to the specifications in Examples 2.1, 2.2, and 2.3.

Example 2.4 (Pairwise beta angular surface) Let

$$h_{\mathbf{x}}(\mathbf{w}) = \frac{\Gamma(3\alpha_{\mathbf{x}}+1)}{\Gamma(2\alpha_{\mathbf{x}}+1)\Gamma(\alpha_{\mathbf{x}})} \sum_{1\le i<j\le 3} h_{i,j_{\mathbf{x}}}(\mathbf{w}),$$

$$h_{i,j_{\mathbf{x}}}(\mathbf{w}) = (w_i+w_j)^{2\alpha_{\mathbf{x}}-1}\left\{1-(w_i+w_j)\right\}^{\alpha_{\mathbf{x}}-1} \times \frac{\Gamma(2\beta_{i,j_{\mathbf{x}}})}{\Gamma^2(\beta_{i,j_{\mathbf{x}}})}\left(\frac{w_i}{w_i+w_j}\right)^{\beta_{i,j_{\mathbf{x}}}-1}\left(\frac{w_j}{w_i+w_j}\right)^{\beta_{i,j_{\mathbf{x}}}-1},$$

where $\mathbf{w} = (w_1, w_2, w_3) \in S_3$ and $\alpha, \beta_{i,j} : \mathcal{X} \subset \mathbb{R}^q \to (0,\infty)$ for $1 \le i < j \le 3$. In Figure 2, we consider the case $\alpha_{\mathbf{x}} = \exp\{\exp(x)\}$, $\beta_{1,2_{\mathbf{x}}} = \exp(x)$, $\beta_{1,3_{\mathbf{x}}} = x+1$, and $\beta_{2,3_{\mathbf{x}}} = x+2$, with $x \in [0.8, 3.3]$. For the different considered values of x, different strengths of global and pairwise dependences can be observed. The mass is concentrated mostly at the center of the simplex due to a large global dependence parameter $\alpha_{\mathbf{x}}$, compared to the pairwise dependence parameters.

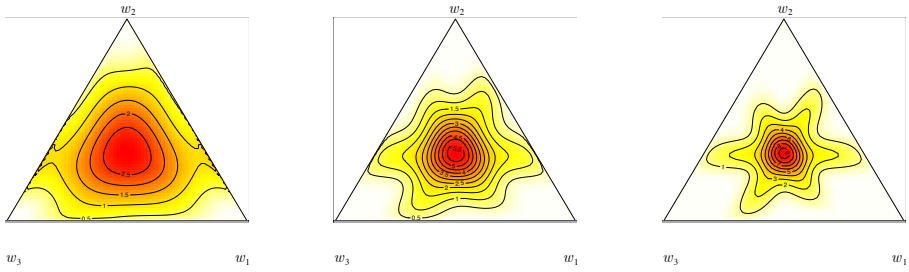

FIGURE 2 Trivariate covariate-adjusted angular density of the pairwise beta model corresponding to the specifications in Example 2.4 with $x = 1.5$ (left), $x = 2.46$ (middle), and $x = 3.22$ (right).

To give the unfamiliar reader insight on some of the quantities introduced in Section 2.2 and more specifically in the VGAM (11), we identify these quantities in the examples mentioned above:

- In Examples 2.1 and 2.3, $d = 2$, $M = 1$, $p = 1$, $q = 1$, and $\mathcal{X} = [0.1, 2]$. The difference between the VGAMs modeled in these two examples resides in the form of dependence of η on x and the link function g. In Example 2.1, the parameter $\theta_x \in (0,1]$, $\eta = x^2 - 0.5x - 1$, and the link function g is the logit function, whereas in Example 2.3 the parameter $\theta_x \in (0,\infty)$, $\eta = x$, and the link function g is the logarithm function.
- In Example 2.2, $d = 2$, $M = 1$, $p = 2$, $q = 1$, $\mathcal{X} = [0.9, 3]$, and $\boldsymbol{\eta} = (x, x)^\top$. The vector of parameters for the bivariate Dirichlet angular density $\boldsymbol{\theta}_x \in (0,\infty)^2$ and the link functions g_1 and g_2 are the logarithm and the square root functions, respectively.
- In Example 2.4, $d = 3$, $M = 2$, $p = 4$, $q = 1$, $\mathcal{X} = [0.8, 3.3]$, and $\boldsymbol{\eta} = (\exp(x), x, \log(x+1), \log(x+2))^\top$. The vector of parameters for the pairwise beta angular density $\boldsymbol{\theta}_x \in (0,\infty)^4$ and the link function g_l is the logarithm function, for $l = 1, \ldots, 4$.

3 | INFERENCE AND ASYMPTOTIC PROPERTIES

The log-likelihood (9) with a covariate-dependent vector of parameters $\boldsymbol{\theta}_{\mathbf{x}}$ is now written as

$$\begin{aligned} \ell(\boldsymbol{\beta}) &:= \sum_{i=1}^{n_{\mathbf{r}}} c_i + \log\left\{h\left(\mathbf{w}_i;\boldsymbol{\beta}\right)\right\}, \\ &= \sum_{i=1}^{n_{\mathbf{r}}} c_i + \log\left(h\left[\mathbf{w}_i;\mathbf{g}^{-1}\left\{\boldsymbol{\eta}(\mathbf{x}^i)\right\}\right]\right), \end{aligned} \tag{13}$$

where c_i is a constant independent of $\boldsymbol{\beta}$ and $\mathbf{g}^{-1}$ is the componentwise inverse of $\mathbf{g}$.

Incorporating a covariate-dependence in the extremal dependence model through a non-linear smooth model adds considerable flexibility in the modelling of the dependence parameter $\boldsymbol{\theta}_{\mathbf{x}}$. The price to pay for this flexibility is reflected in the estimation procedure. The estimation of $\boldsymbol{\theta}_{\mathbf{x}}$, hence of $\boldsymbol{\beta}$, is performed by maximizing the penalized log-likelihood

$$\ell(\boldsymbol{\beta},\boldsymbol{\gamma}) = \ell(\boldsymbol{\beta}) - \frac{1}{2}\mathbf{J}(\boldsymbol{\gamma}), \tag{14}$$

where the penalty term can be written as

$$\mathbf{J}(\boldsymbol{\gamma}) = \sum_{k=1}^{q} \boldsymbol{\beta}_{[k]}^{\top}\left\{\mathbf{P}_k \otimes \operatorname{diag}(\gamma_{(1)k},\ldots,\gamma_{(p)k})\right\}\boldsymbol{\beta}_{[k]} = \boldsymbol{\beta}^{\top}\mathbf{P}(\boldsymbol{\gamma})\boldsymbol{\beta},$$

with $\mathbf{P}(\boldsymbol{\gamma})$ a $p(1+q\tilde{d}) \times p(1+q\tilde{d})$ block matrix with a first $p \times p$ block filled with zeros and q blocks, each formed by a $p\tilde{d} \times p\tilde{d}$ matrix $\mathbf{P}_k$ that depends only on the knots of the B-spline functions for the covariate $\mathbf{x}_k$. The matrix $\mathbf{P}(\boldsymbol{\gamma})$ can be written as $\mathbf{P}(\boldsymbol{\gamma}) = \tilde{\mathbf{X}}^{\top}\tilde{\mathbf{X}}$ for some $p(1+q\tilde{d}) \times p(1+q\tilde{d})$ real matrix $\tilde{\mathbf{X}}$. The vectors $\boldsymbol{\beta}_{[k]}$ are defined in (12), and $\gamma_{(l)k}$ are termed the smoothing parameters.

The penalty term in (14) controls the wiggliness and the fidelity to the data of the component functions in (11) through the vector $\boldsymbol{\gamma}$ of the smoothing parameters $\gamma_{(l)k}$ for $l = 1,\ldots,p$ and $k = 1,\ldots,q$. Larger values of $\gamma_{(l)k}$ lead to smoother effects of the covariate $\mathbf{x}_k$ on the lth component of $\boldsymbol{\eta}$.

The maximization of the penalized log-likelihood (14) is based on a Newton–Raphson (N–R) algorithm. At each step of the N–R algorithm, a set of smoothing parameters is proposed by outer iteration [Wood, 2017], and a penalized iterative reweighted least squares (PIRLS) algorithm is performed, in an inner iteration, to update the model coefficients estimates. We detail the inner fitting procedure in the following section and the outer iteration in Section 3.2.

3.1 Fitting algorithm

We suppose that the penalized log-likelihood (14) depends only on the $p(1+q\tilde{d})$-vector $\boldsymbol{\beta}$ and that the vector of smoothing parameters $\boldsymbol{\gamma}$ is proposed (at each iteration of the N–R algorithm) by outer iteration and is therefore fixed in what follows.

The penalized maximum log-likelihood estimator (PMLE) $\hat{\boldsymbol{\beta}}$ satisfies the following score equation

$$\frac{\partial\ell(\hat{\boldsymbol{\beta}},\boldsymbol{\gamma})}{\partial\boldsymbol{\beta}} = \mathbf{X}_{\mathrm{VAM}}^{\top}\mathbf{u}(\hat{\boldsymbol{\beta}}) - \mathbf{P}(\boldsymbol{\gamma})\hat{\boldsymbol{\beta}} = \mathbf{0},$$

where $\mathbf{u}(\boldsymbol{\beta}) = \partial\ell(\boldsymbol{\beta})/\partial\boldsymbol{\eta} \in \mathbb{R}^{pn_{\mathbf{r}}}$ and $\mathbf{X}_{\mathrm{VAM}}$ is as defined in (12). To obtain $\hat{\boldsymbol{\beta}}$, we update $\boldsymbol{\beta}^{(a-1)}$, the $(a-1)$th estimate

of the true $\boldsymbol{\beta}_0$, by Newton–Raphson:

$$\boldsymbol{\beta}^{(a)} = \boldsymbol{\beta}^{(a-1)} + \mathbf{I}\left(\boldsymbol{\beta}^{(a-1)}\right)^{-1}\left\{\mathbf{X}_{\text{VAM}}^{\top}\mathbf{u}(\boldsymbol{\beta}^{(a-1)}) - \mathbf{P}(\boldsymbol{\gamma})\boldsymbol{\beta}^{(a-1)}\right\}, \tag{15}$$

where

$$\begin{cases} \mathbf{I}\left(\boldsymbol{\beta}^{(a-1)}\right) = -\dfrac{\partial^2 \ell(\boldsymbol{\beta},\boldsymbol{\gamma})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\top}} = \mathbf{X}_{\text{VAM}}^{\top}\mathbf{W}(\boldsymbol{\beta}^{(a-1)})\mathbf{X}_{\text{VAM}} + \mathbf{P}(\boldsymbol{\gamma}), \\ \mathbf{W}(\boldsymbol{\beta}^{(a-1)}) = -\dfrac{\partial^2 \ell(\boldsymbol{\beta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^{\top}} \in \mathbb{R}^{pn_{\mathbf{r}}\times pn_{\mathbf{r}}}. \end{cases}$$

The matrix $\mathbf{W}(\boldsymbol{\beta}^{(a-1)})$ is termed the working weight matrix. If the expectation $\mathrm{E}\{\partial^2\ell(\boldsymbol{\beta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^{\top}\}$ is obtainable, a Fisher scoring algorithm is then preferred, as it ensures the positive definiteness of $\mathbf{W}(\boldsymbol{\beta})$ over a larger region of the parameter space $\mathbf{B}$ than in the N–R algorithm. When the working weight matrix is not positive definite, which might happen when the parameter $\boldsymbol{\beta}^{(a-1)}$ is far from the true $\boldsymbol{\beta}_0$, a Greenstadt [Greenstadt, 1967] modification is applied, and the negative eigenvalues of $\mathbf{W}(\boldsymbol{\beta}^{(a-1)})$ are replaced by their absolute values. With the different families of angular densities considered in Examples 2.1, 2.2, and 2.4, the expected information matrix is not obtainable and is hence replaced by the observed information matrix on which a Greenstadt modification is applied whenever needed. See Yee [2015, Section 9.2] for other remedies and techniques for deriving well-defined working weight matrices.

Let $\mathbf{z}^{(a-1)} := \mathbf{X}_{\text{VAM}}\boldsymbol{\beta}^{(a-1)} + \mathbf{W}(\boldsymbol{\beta}^{(a-1)})^{-1}\mathbf{u}(\boldsymbol{\beta}^{(a-1)})$ be the $pn_{\mathbf{r}}$-vector of working responses. Then, (15) can be rewritten in a PIRLS form as

$$\begin{aligned} \boldsymbol{\beta}^{(a)} &= \left\{\mathbf{X}_{\text{VAM}}^{\top}\mathbf{W}(\boldsymbol{\beta}^{(a-1)})\mathbf{X}_{\text{VAM}} + \mathbf{P}(\boldsymbol{\gamma})\right\}^{-1}\mathbf{X}_{\text{VAM}}^{\top}\mathbf{W}(\boldsymbol{\beta}^{(a-1)})\mathbf{z}^{(a-1)} \\ &= \left\{\mathbf{X}_{\text{PVAM}}^{\top}\tilde{\mathbf{W}}^{(a-1)}\mathbf{X}_{\text{PVAM}}\right\}^{-1}\mathbf{X}_{\text{PVAM}}^{\top}\tilde{\mathbf{W}}^{(a-1)}\mathbf{y}^{(a-1)}, \end{aligned}$$

where $\mathbf{X}_{\text{PVAM}}$, $\mathbf{y}^{(a-1)}$, and $\tilde{\mathbf{W}}^{(a-1)}$ are augmented versions of $\mathbf{X}_{\text{VAM}}$, $\mathbf{z}^{(a-1)}$ and $\mathbf{W}(\boldsymbol{\beta}^{(a-1)})$, respectively, and are defined as

$$\begin{cases} \mathbf{X}_{\text{PVAM}} = \begin{pmatrix} \mathbf{X}_{\text{VAM}}^{\top} & \tilde{\mathbf{X}} \end{pmatrix}^{\top} \in \mathbb{R}^{p(1+n_{\mathbf{r}}+q\tilde{d})\times p(1+q\tilde{d})}, \\ \mathbf{y}^{(a-1)} = \begin{pmatrix} \mathbf{z}^{(a-1)} & \mathbf{0}_{p(1+q\tilde{d})} \end{pmatrix}^{\top} \in \mathbb{R}^{p(1+n_{\mathbf{r}}+q\tilde{d})}, \\ \tilde{\mathbf{W}}^{(a-1)} = \operatorname{diag}\left(\mathbf{W}(\boldsymbol{\beta}^{(a-1)}), \mathbf{I}_{p(1+q\tilde{d})\times p(1+q\tilde{d})}\right) \in \mathbb{R}^{p(1+n_{\mathbf{r}}+q\tilde{d})\times p(1+n_{\mathbf{r}}+q\tilde{d})}. \end{cases}$$

The algorithm stops when the change in the coefficients $\boldsymbol{\beta}$ between two successive iterations is sufficiently small. Convergence of the N–R algorithm is not guaranteed and might not occur if the quadratic approximation of $\ell(\boldsymbol{\beta},\boldsymbol{\gamma})$ around $\hat{\boldsymbol{\beta}}$ is poor. See Yee [2015, 2016] for more details.

The plug-in penalized maximum log-likelihood estimator of the covariate-dependent angular density is defined as

$$\widehat{h}_{\mathbf{x}}(\mathbf{w}) \equiv h\{\mathbf{w}; \mathbf{g}^{-1}(\mathbf{X}_{\text{VAM}}\hat{\boldsymbol{\beta}})\}. \tag{16}$$

In the following section, we give details about the selection of the smoothing parameters $\boldsymbol{\gamma}$, which is outer to the PIRLS algorithm.

3.2 Selection of the smoothing parameters

To implement the PIRLS algorithm performed at each iteration of the N–R algorithm, a smoothing parameter selection procedure is conducted by minimizing a prediction error estimate given by the generalized cross validation (GCV) score. Let $\mathbf{A}^{(a-1)}(\boldsymbol{\gamma})$ be the influence matrix of the fitting problem at the ath iteration, defined as

$$\mathbf{A}^{(a-1)}(\boldsymbol{\gamma}) = \mathbf{X}_{\text{PVAM}} \left\{ \mathbf{X}_{\text{PVAM}}^{\top} \tilde{\mathbf{W}}^{(a-1)} \mathbf{X}_{\text{PVAM}} \right\}^{-1} \mathbf{X}_{\text{PVAM}}^{\top} \tilde{\mathbf{W}}^{(a-1)}.$$

Then, by minimizing the GCV score

$$\text{GCV}^{(a-1)} = \frac{n_{\mathbf{r}} \left\{ \mathbf{y}^{(a-1)} - \mathbf{A}^{(a-1)}(\boldsymbol{\gamma}) \mathbf{y}^{(a-1)} \right\}^{\top} \tilde{\mathbf{W}}^{(a-1)} \left\{ \mathbf{y}^{(a-1)} - \mathbf{A}^{(a-1)}(\boldsymbol{\gamma}) \mathbf{y}^{(a-1)} \right\}}{\left[n_{\mathbf{r}} - \text{trace} \left\{ \mathbf{A}^{(a-1)}(\boldsymbol{\gamma}) \right\} \right]^2},$$

we aim at balancing between goodness of fit and complexity of the model, which is measured by the trace of the influence matrix and termed the effective degrees of freedom (EDF). The EDF of the fitted VGAM (12) are defined as the EDF obtained at convergence, that is, trace $\left\{ \mathbf{A}^{(c-1)}(\boldsymbol{\gamma}) \right\}$, where c is the iteration at which convergence occurs.
Both the fitting algorithm of Section 3.1 and the smoothing parameter selection are implemented in the `R` package `VGAM` [Yee, 2017], with the latter being required from the `R` package `mgcv` [Wood, 2017].
Model selection between different, not necessarily nested, fitted VGAMs is performed based on the Akaike information criterion (AIC), where the number of parameters of the model is replaced by its EDF to account for penalization. More details on the (conditional) AIC for models with smoothers along with a corrected version of this criterion, which takes into account the smoothing parameter uncertainty, can be found in Wood [2017, Section 6.11].

3.3 Large sample properties

We now derive the consistency and asymptotic normality of the PMLE $\hat{\boldsymbol{\beta}}$ defined in Section 3.1. Throughout this section, we assume that the set of angular observations $\{\mathbf{w}_i\}_{i=1}^{n_{\mathbf{r}}}$ defined in the log-likelihood (13), stems from a parametric family of angular densities of multivariate extreme value distributions $\{h(\cdot, \boldsymbol{\beta}), \boldsymbol{\beta} \in \mathbf{B}\}$, with $\mathbf{B}$ a compact subset of $\mathbb{R}^{p(1+q\tilde{d})}$. The following large sample properties are derived for an increasing size $n_{\mathbf{r}}$ and under the assumption that the angular density $h(\cdot; \boldsymbol{\beta})$ has all its density in the interior of the unit simplex S_d. Additionally, we assume that the dependence parameter of the angular density is a linear combination of known spline functions where the number and the location of the internal knots are fixed, that is, the dimension of the parameter $\boldsymbol{\beta}$ is fixed. The bias resulting from this assumption is negligible compared to the bias due to the smoothing parameter uncertainty [Ruppert, 2002].
Based on the penalized log-likelihood (14), $\hat{\boldsymbol{\beta}}$ satisfies the following score equation

$$\mathbf{m}(\boldsymbol{\beta}) - \mathbf{P}(\boldsymbol{\gamma})\boldsymbol{\beta} = \mathbf{0}_{p(1+q\tilde{d})}, \tag{17}$$

where $\mathbf{m}(\boldsymbol{\beta}) = \partial \ell(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$.
Let $\mathbf{B}_0$ be an open neighbourhood around the true parameter $\boldsymbol{\beta}_0$. Moreover, we define $\mathbf{m}(\mathbf{w}, \boldsymbol{\beta}) = \partial \log\{h(\mathbf{w}; \boldsymbol{\beta})\} / \partial \boldsymbol{\beta}$.
Our asymptotic results hold under the following customary assumptions on the smoothing parameters $\boldsymbol{\gamma}$ and the angular density $h(\cdot; \boldsymbol{\beta})$:

(A1) $\boldsymbol{\gamma} = \begin{pmatrix} \gamma_{(1)1} & \cdots & \gamma_{(p)1} & \cdots & \gamma_{(1)q} & \cdots & \gamma_{(p)q} \end{pmatrix}^{\top} = o(n_{\mathbf{r}}^{1/2}) 1_{pq}$.

(A2) Regularity conditions:

- The angular density support $\{\mathbf{w} \in S_d : h(\mathbf{w};\boldsymbol{\beta}) > 0\}$ does not depend on $\boldsymbol{\beta} \in \mathbf{B}_0$.
- If $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, then $h(\mathbf{w};\boldsymbol{\beta}) \neq h(\mathbf{w};\boldsymbol{\beta}_0)$, with $\boldsymbol{\beta} \in \mathbf{B}_0$. Moreover, $\mathrm{E}[\sup_{\boldsymbol{\beta}\in\mathbf{B}} |\log\{h(\mathbf{w};\boldsymbol{\beta})\}|] < \infty$.
- For $\mathbf{w} \in S_d$, $h(\mathbf{w};\boldsymbol{\beta}) \in C^3(\mathbf{B})$ and $h(\mathbf{w};\boldsymbol{\beta}) > 0$ on $\mathbf{B}_0$.
- $\int \sup_{\boldsymbol{\beta}\in\mathbf{B}} \|\mathbf{m}(\mathbf{w},\boldsymbol{\beta})\| \, d\mathbf{w} < \infty$ and $\int \sup_{\boldsymbol{\beta}\in\mathbf{B}} \|\partial\mathbf{m}(\mathbf{w},\boldsymbol{\beta})/\partial\boldsymbol{\beta}^\top\| d\mathbf{w} < \infty$.
- For $\boldsymbol{\beta} \in \mathbf{B}_0$, $\mathbf{i}(\boldsymbol{\beta}) := \mathrm{cov}\{\mathbf{m}(\mathbf{W},\boldsymbol{\beta})\} = \mathbf{X}_{\mathrm{VAM}}^\top \mathbf{W}(\boldsymbol{\beta}) \mathbf{X}_{\mathrm{VAM}}$ exists and is positive-definite.
- For each triplet $1 \leq q, r, s \leq p(1 + q\tilde{d})$, there exists a function $M_{qrs} : S_d \to \mathbb{R}$ such that, for $\mathbf{w} \in S_d$ and $\boldsymbol{\beta} \in \mathbf{B}_0$, $|\partial^3 \log\{h(\mathbf{w};\boldsymbol{\beta})\}/\partial\beta_{qrs}| \leq M_{qrs}(\mathbf{w})$, and $\mathrm{E}\left\{M_{qrs}(\mathbf{W})\right\} < \infty$.

Assumption (A1) is needed to control the influence of the smoothing parameters as $n_{\mathbf{r}} \to \infty$ such that the asymptotic unbiasedness of $\hat{\boldsymbol{\beta}}$ can be established. This assumption is rather weak as it allows the smoothing parameters to grow as the size of the threshold exceedances grows, at a rate smaller than $n_{\mathbf{r}}^{1/2}$, implying therefore that heavy oversmoothing is avoided. As discussed by Marra and Wood [2012], heavy oversmoothing is likely to happen under two situations. The first one corresponds to situations where two covariates are highly correlated but the effect of one is very smooth and the effect of the other is very wiggly, leading therefore to a possible inversion of the degrees of smoothness and to one of the covariates being highly oversmoothed. The second situation corresponds to the setting where the true effect of one of the covariates is close to a function in the null space of the penalty associated to this covariate, i.e., a straight line, and might therefore be estimated exactly as this function. Assumption (A2) consists of standard regularity assumptions under which the classical asymptotic properties of MLEs hold. These assumptions are expressed in terms of the angular density as the first term in the log-likelihood (9) does not influence the maximization. Although difficult to verify, these assumptions are similar to the ones considered in Bienvenüe and Robert [2017] and Padoan et al. [2010] where in the latter, the assumptions are imposed on the components of the composite likelihood along with additional assumptions on the composite score equation.
The next theorem characterizes the large sample behaviour of our estimator $\hat{\boldsymbol{\beta}}$.

Theorem 1 *Let $\{h(\cdot,\boldsymbol{\beta}), \boldsymbol{\beta} \in \mathbf{B}\}$ be a parametric family of angular densities of multivariate extreme value distributions, where $\mathbf{B}$ is a compact subset of $\mathbb{R}^{p(1+q\tilde{d})}$. Let $\boldsymbol{\beta}_0$ be an interior point of $\mathbf{B}$ and $\mathbf{B}_0$ an open neighbourhood around $\boldsymbol{\beta}_0$. Under (A1) and (A2), and for independent observations $\mathbf{w}_1, \ldots, \mathbf{w}_{n_{\mathbf{r}}}$ with distribution $h(\cdot,\boldsymbol{\beta}_0)$, it follows that, as $n_{\mathbf{r}} \to \infty$, the estimator $\hat{\boldsymbol{\beta}}$ maximizing the penalized log-likelihood* (14) *verifies:*

1. $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(n_{\mathbf{r}}^{-1/2})$.
2. $n_{\mathbf{r}}^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathrm{d}} N(0, \mathbf{i}(\boldsymbol{\beta}_0)^{-1})$.

These results are derived from a second-order Taylor expansion of the score equation (17) around the true parameter $\boldsymbol{\beta}_0$ along the same lines as in Vatter and Chavez-Demoulin [2015] and Davison [2003, p. 147]. The proof of Theorem 1 is deferred to the Supplementary Materials. Similar results on the large sample behaviour of the corresponding plug-in estimator (16) can be derived using the multivariate delta method. These results are useful to derive and construct approximate confidence intervals for conditional angular densities and to compare nested models based on likelihood ratio tests. Our proviso is similar to that of de Carvalho and Davison [2014] in the sense that asymptotic properties of the estimator $\hat{\boldsymbol{\beta}}$ are derived under the assumption of known margins and we sample from the limiting object $h_{\mathbf{x}}$, whereas in practice only a sample of (estimated) pseudo-angles, $\{\widehat{\mathbf{w}}_i\}_{i=1}^{n_{\mathbf{r}}}$, would be available and the uncertainty arising from the marginal fitting would not be accounted for. Asymptotic properties under misspecification of the parametric model set for $h_{\mathbf{x}}$ could in principle be derived under additional assumptions on $\boldsymbol{\beta}$ and $\mathbf{m}$, along the same lines as in

standard likelihood theory [Knight, 2000]. The resulting theory is outside the scope of this work and is deliberately not studied here.

4 | SIMULATION STUDY

4.1 Data generating processes and preliminary experiments

We assess the performance of our methods using the bivariate extremal dependence structures presented in Section 2.2—and displayed in Figure 1—as well as the trivariate pairwise beta dependence model from Example 2.4—depicted in Figure 2. Monte Carlo evidence will be reported in Section 4.2. For now, we concentrate on illustrating the methods over a single-run experiment on these scenarios. For each dependence model from Examples 2.1–2.3, we draw a sample $\{(w_{i,1}, w_{i,2})\}_{i=1}^{n_{\mathbf{r}}}$ from the corresponding angular density h_x with sample size $n_{\mathbf{r}} = 300$ and where each angular observation $(w_{i,1}, w_{i,2})$ is drawn from the chosen dependence model conditional on a fixed value x^i of the covariate x. To gain insight into the bias and variance of our covariate-adjusted spectral density estimator, we compute its 95% asymptotic confidence bands based on Theorem 1 and at different values of w in $(0, 1)$. The only source of bias in our estimation procedure is due to the penalization of the model likelihood causing a smoothing bias [Wood, 2017] if the smoothing parameters do not vanish at a certain rate (see Section 3.3). The uncertainty due to the choice of the parametric model is deliberately not taken into account, that is, the simulations are performed in a well-specified framework.

Figure 3 displays the estimates of the covariate-adjusted spectral densities from Examples 2.1, 2.2, and 2.3 for various fixed values of the covariate x that induce different extremal dependence strengths. All panels show that for the different extremal dependence schemes (strength and asymmetry), the covariate-adjusted spectral densities are accurately estimated and the true curves fall well within the 95% confidence bands. The estimates in the Dirichlet case seem to be a bit more biased, and this might be explained by the fact that both of the two non-orthogonal parameters of the model depend smoothly on the covariate x.

We now consider the case of the trivariate pairwise beta dependence model from Example 2.4. We draw a sample $\{(w_{i,1}, w_{i,2}, w_{i,3})\}_{i=1}^{n_{\mathbf{r}}}$ with sample size $n_{\mathbf{r}} = 300$ where each observation $(w_{i,1}, w_{i,2}, w_{i,3})$ is drawn from the pairwise beta model conditional on a fixed value x_i of the covariate x, as illustrated in Figure 2. Figure 4 displays the contour plots of the estimates of the covariate-adjusted spectral density from Example 2.4 at three fixed values of x. All panels in Figure 4 show that, for the different extremal dependence schemes, i.e., for the different considered values of x, the contour plots of the estimates are remarkably close to the actual contour plots. The estimates are slightly more biased near the edges of the simplex than in the center, reflecting a better estimation of the global dependence parameter compared to the pairwise dependence parameters.

4.2 Monte Carlo evidence

A Monte Carlo study was conducted by simulating 500 independent samples of sizes $n_{\mathbf{r}} = 300$ and $n_{\mathbf{r}} = 500$ angular observations, respectively. In what follows we focus on documenting how the level of accuracy increases when the number of observations increases by assessing the mean integrated absolute error (MIAE)—which for the bivariate case can be written as

$$\mathrm{MIAE} = \mathrm{E}\left\{\int_{\mathcal{X}}\int_0^1 |\widehat{h}_x(w) - h_x(w)|\,\mathrm{d}w\,\mathrm{d}x\right\}.$$

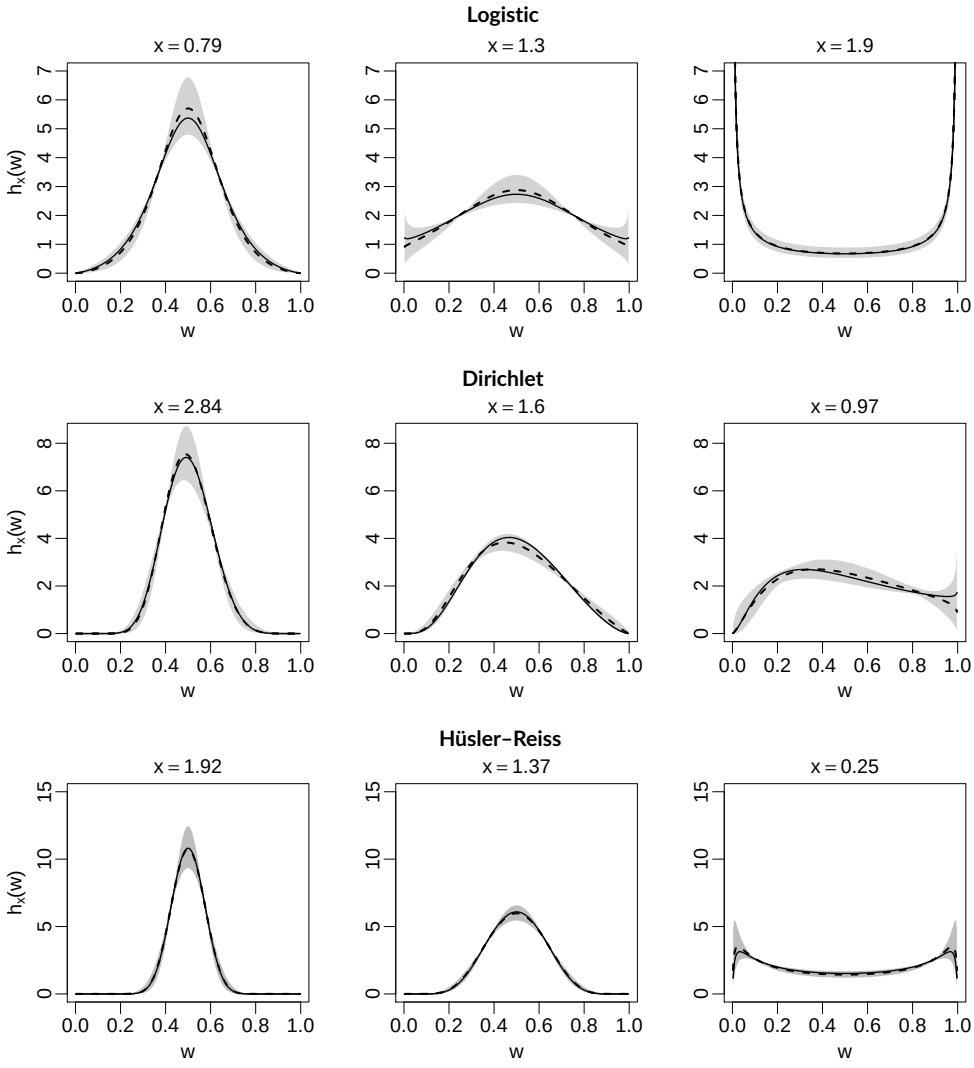

FIGURE 3 Estimates of the covariate-adjusted spectral densities in Examples 2.1, 2.2, and 2.3 conditional on different values of the covariate x (dashed lines) along with their 95% (pointwise) asymptotic confidence bands (grey area). The true spectral densities are displayed in solid lines.

The results are reported in Table 1.

As expected, an increase in the number of angular observations leads to a reduction of MIAE. To give a more granular level of detail than that of Table 1 on the behaviour of the estimator over specific values of the covariate and of the unit simplex, Figure 5 displays the Monte Carlo confidence intervals of the covariate-adjusted spectral densities from Examples 2.1–2.3 for various fixed values of the covariate x, along with the Monte Carlo means and Figure 6 displays the contour plots of the Monte Carlo mean of the covariate-adjusted spectral density from Example 2.4 at three fixed values of x.

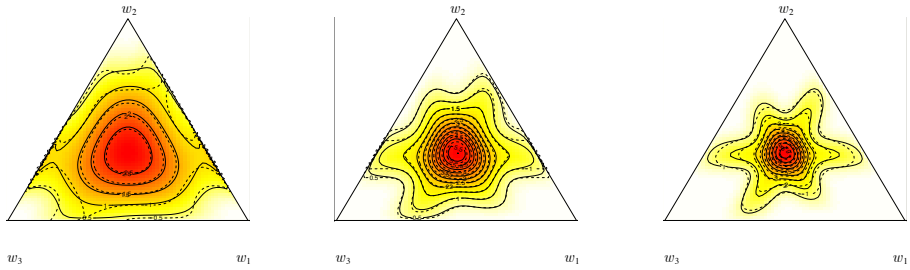

FIGURE 4 Contour plots of the covariate-adjusted pairwise beta spectral density estimate (dashed lines) at $x = 1.5$ (left), $x = 2.46$ (middle), and $x = 3.22$ (right). The contour plots of the true spectral density are displayed in solid lines.

TABLE 1 Mean integrated absolute error (MIAE) estimates computed from 500 samples for the covariate-adjusted spectral densities in Examples 2.1–2.3; $n_{\mathbf{r}}$ denotes the number of angular observations

$n_{\mathbf{r}}$	Covariate-adjusted angular density	MIAE
	Logistic	0.2185
300	Dirichlet	0.3372
	Hüsler–Reiss	0.2205
	Logistic	0.1781
500	Dirichlet	0.2459
	Hüsler–Reiss	0.1648

As can be seen from Figures 5 and 6, our method successfully recovers the corresponding target covariate-adjusted angular densities with a high level of precision over the simulation study. Additionally, the variability in the Monte Carlo study in Figure 5 is comparable to the asymptotic variability displayed in Figure 3. Finally, an assessment of the sensitivity of our method to the choice of the radial threshold is conducted in the two-parameter Dirichlet example 2.2 where a sample $\{(Y_1^i, Y_2^i)\}_{i=1}^n$ is drawn from the corresponding bivariate extreme value distribution G_x. Each observation (Y_1^i, Y_2^i) has unit Fréchet margins and is drawn from the chosen dependence model conditional on a fixed value x^i of the covariate x. For estimating h_x, we fix the sample size at $n = 6000$ and consider the observations with a radial component exceeding its quantile at the level 93%, 95%, and 97%, ending up with $n_{\mathbf{r}} = 420, 300, 180$ extreme (angular) observations, respectively. Figure 7 displays, for the different considered radial thresholds, the Monte Carlo confidence intervals of the limiting covariate-adjusted spectral density from Example 2.2 for various fixed values of the covariate x, along with the Monte Carlo means.

Although the angular observations resulting from thresholding the radial component do not come from the true angular density that we model, our method recovers relatively accurately the true limiting angular density. Comparing the results of Figure 7 with the middle panels of Figure 5 where simulation is performed from the true angular density, we observe as expected a slight bias due to the sub-asymptotic modelling of angular observations, but which decreases

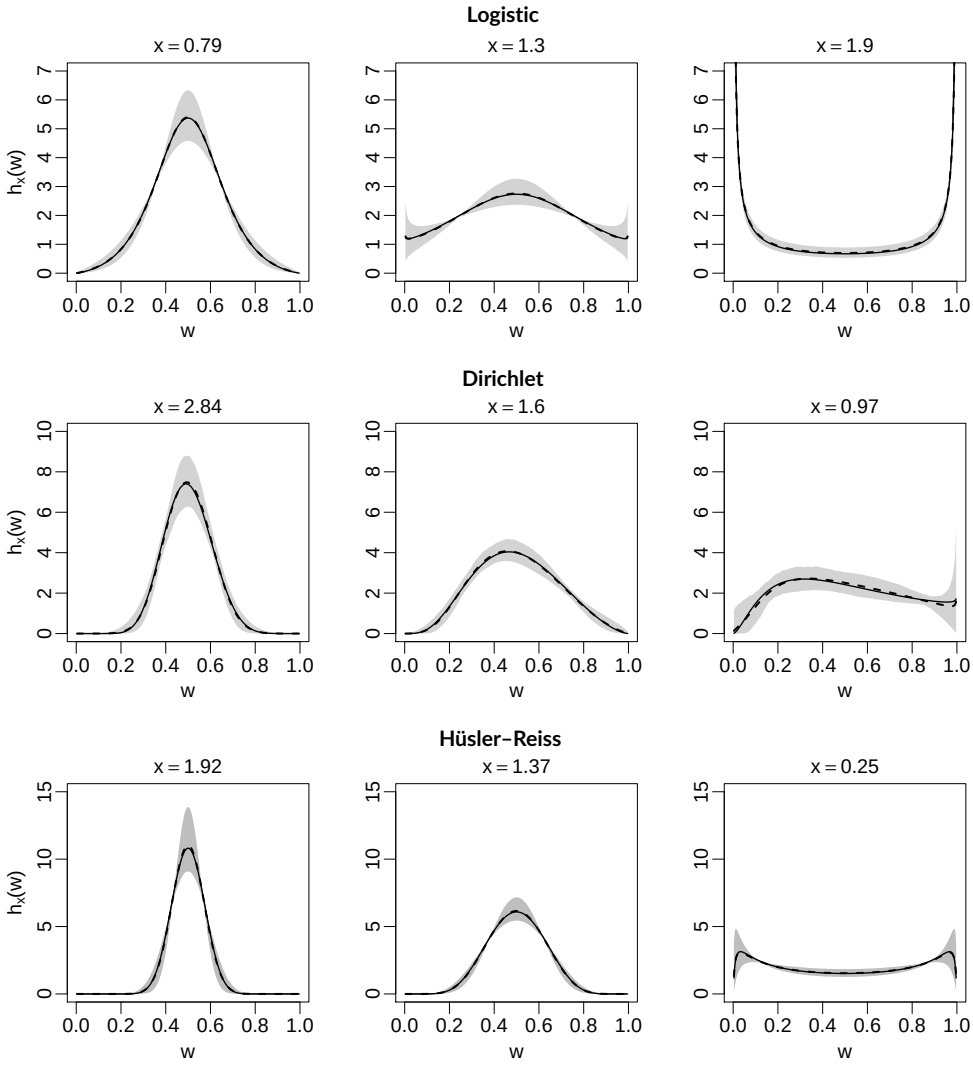

FIGURE 5 The Monte Carlo 95% confidence intervals of the spectral densities in Examples 2.1, 2.2, and 2.3 conditional on different values of the covariate x (grey area) along with their Monte Carlo means (dashed lines). The true spectral densities are displayed in solid lines.

as the radial threshold increases. The bias is more pronounced for a value of the covariate x inducing weak asymptotic dependence due to the residual dependence that vanishes asymptotically but is observed at finite thresholds.

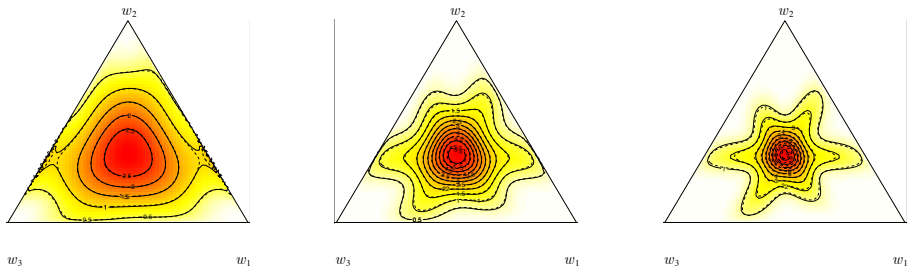

FIGURE 6 Contour plots of the Monte Carlo mean estimate of the covariate-adjusted pairwise beta spectral density (dashed lines) at $x = 1.5$ (left), $x = 2.46$ (middle), and $x = 3.22$ (right). The contour plots of the true spectral density are displayed in solid lines.

5 | EXTREME TEMPERATURE ANALYSIS

5.1 Data description, motivation for the analysis, and preprocessing

In this section, we describe an application to modelling the dependence between extreme air winter (December–January–February) temperatures at two sites in the Swiss Alps: Montana—at an elevation of 1427m—and Zermatt—at an elevation of 1638m. The sites are approximatively 37km apart.

In the Alpine regions of Switzerland, there is an obvious motivation to focus on extreme climatic events, as their impact on the local population and infrastructure can be very costly. As stated by Beniston [2007], warm winter spells, that is, periods with strong positive temperature exceedances in winter, can exert significant impacts on the natural ecosystems, agriculture, and water supply:

> *"Temperatures persistently above 0°C will result in early snow-melt and a shorter seasonal snow cover, early water runoff into river basins, an early start of the vegetation cycle, reduced income for alpine ski resorts and changes in hydro-power supply because of seasonal shifts in the filling of dams [Beniston, 2004]."*

In this analysis, we are interested in the dynamics of the dependence between extreme air temperatures in Montana and Zermatt during the winter season. The dynamics of both extreme high and extreme low winter temperatures in these two sites will be assessed and linked to the following explanatory factors: time (in years) (t), day within season (d_s), and the NAO (North Atlantic Oscillation) index (z); the latter is a normalized pressure difference between Iceland and the Azores that is known to have a major direct influence on the alpine region temperatures, especially during winter [Beniston, 2005]. The choice of the studied sites is of great importance in this analysis. Beniston and Rebetez [1996] showed that both cold and warm winters exhibit temperature anomalies that are altitude-dependent, with high-elevation resorts being more representative of free atmospheric conditions and less likely to be contaminated by urban effects. Therefore, to study the "pure" effect of the above-mentioned explanatory covariates on the winter temperature extremal dependence, we choose the two high elevation sites Montana and Zermatt.

The data consist of daily winter temperature minima and maxima measured at 2m above ground surface and were obtained from the MeteoSwiss website

www.meteoswiss.admin.ch

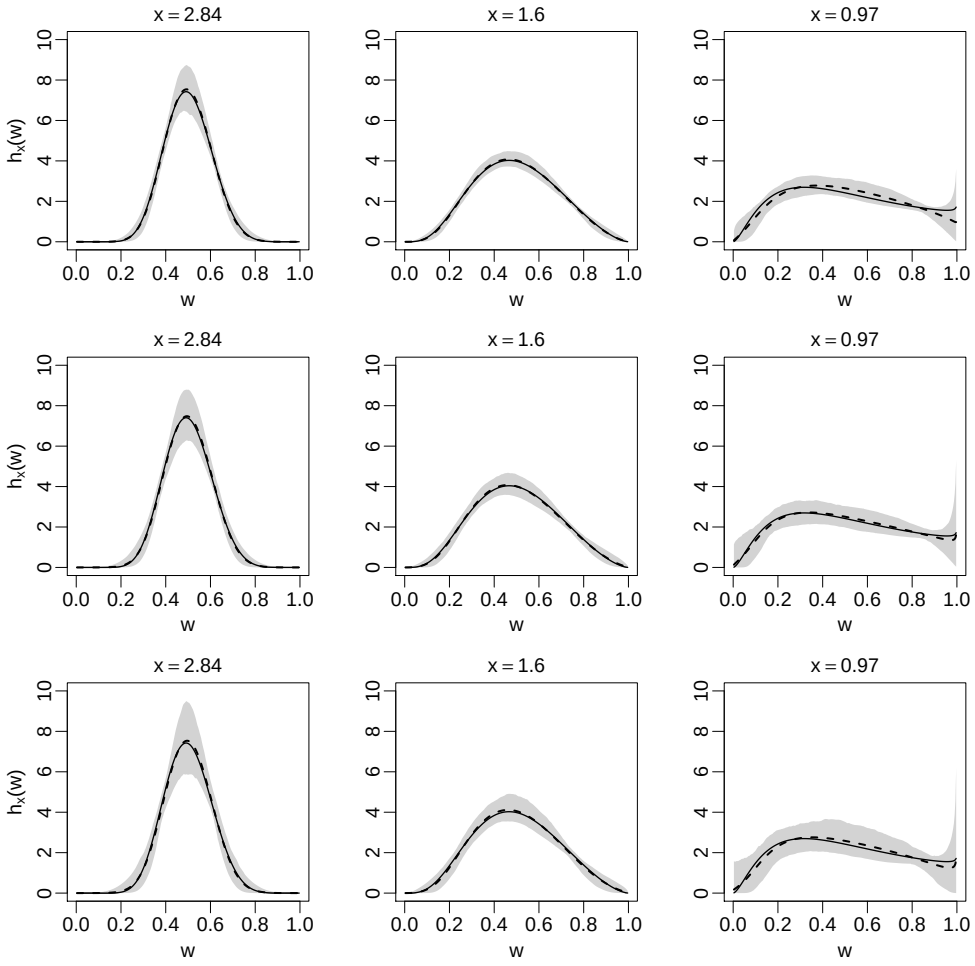

FIGURE 7 The Monte Carlo 95% confidence intervals of the limiting spectral density in Example 2.2 conditional on different values of the covariate x (grey area) along with their Monte Carlo means (dashed lines). Radial thresholds at the 93% quantile level (top panels), the 95% quantile level (middle panels), and the 97% quantile level (bottom panels) are considered. The true limiting spectral density is displayed in solid lines.

The data were available from 1981 to 2016, giving a total of 3190 winter observations per site. Daily NAO index measurements were obtained from the NOAA (National Centers for Environmental Information), at

`www.ngdc.noaa.gov/ftp.html`

We first transform the minimum temperature data by multiplication by -1 and then fit at each site—and to both daily minimum and maximum temperatures—a generalized Pareto Distribution (GPD) [Coles, 2001, ch. 4]

$$G_{\sigma,\xi}(y) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)_+^{-1/\xi}, \tag{18}$$

to model events above the 95% quantile u_{95} for each of the four temperature time series. In (18), $\sigma > 0$ is the scale parameter that depends on u_{95}, and $-\infty < \xi < \infty$ is the shape parameter. As is common with temperature data analysis, we test the effect of the considered covariates on the behaviour of the threshold exceedances by allowing the scale parameter of the GPD (18) to smoothly vary with t, z, and d_s [Chavez-Demoulin and Davison, 2005]. Model selection is based on the double penalty method of Marra and Wood [2011] allowing non-significant covariates to be shrunk out of the model. Except for the day in season d_s which was found to have a significant effect on the scale parameter for the threshold exceedances of the Zermatt daily minimum temperatures, the considered covariates were found to be non-significant with positive but very small effective degrees of freedom (EDF). Graphical goodness-of-fit tests for the four resulting GPD models are conducted by comparing the distribution of a test statistic S with the unit exponential distribution (if $Y \sim G_{\sigma,\xi}$, then $S = -\ln\{1 - G_{\sigma,\xi}(Y)\}$ is unit exponentially distributed). Figure 8 displays the resulting qq-plots and confirms the validity of these models. The fitted models are then used to transform the

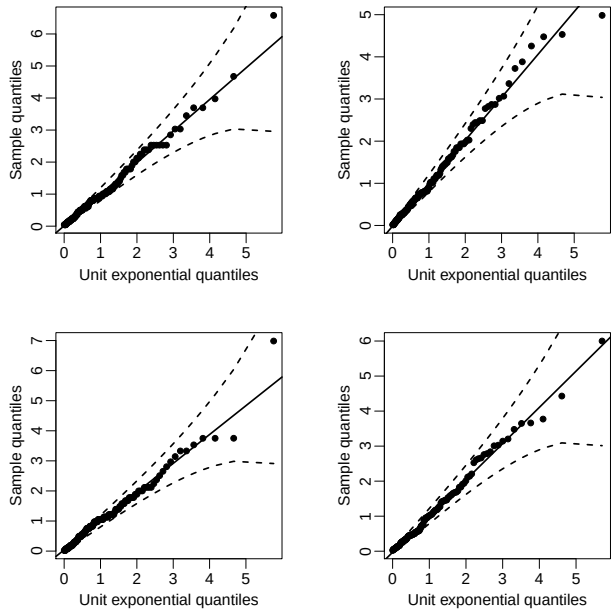

FIGURE 8 Diagnostic plots of the GPD modelling of the threshold exceedances of the daily maximum winter temperatures (left) and the daily minimum winter temperatures (right) in Montana (top) and Zermatt (bottom).

data to a common unit Fréchet scale by probability integral transform and where the empirical distribution is used below u_{95}. This results in two datasets of bivariate observations (in Montana and Zermatt) with unit Fréchet margins: one for the daily maximum temperatures and the other one for the daily minimum temperatures.

Following the theory developed in Section 2.1, we transform each of the two datasets into pseudo-datasets of radial and angular components. By retaining the angular observations corresponding to a radial component exceeding its 95% quantile in each pseudo-dataset, we end up with two pseudo-samples of 160 extreme bivariate (angular) observations in each pseudo-dataset.

5.2 Covariate-adjusted dependence of extreme temperatures

In the following analyses of the dynamics of the dependence between extreme temperatures in Montana and Zermatt—and in line with findings from previous analyses of extreme temperatures in Switzerland [Davison and Gholamrezaee, 2011, Davison et al., 2013, Dombry et al., 2013]—we assume asymptotic dependence in both extremely high and extremely low winter temperatures.

5.2.1 Dependence of extreme high winter temperatures

The covariate-adjusted bivariate angular densities presented in Section 2.2 are now fitted to the pseudo-sample of extreme high temperatures. The effects of the explanatory covariates t, z, and d_s are tested in each of the three angular densities: the logistic model (Example 2.1) with parameter $\alpha(t, z, d_s)$, the Dirichlet model (Example 2.2) with parameters $\alpha(t, z, d_s)$ and $\beta(t, z, d_s)$, and the Hüsler–Reiss model (Example 2.3) with parameter $\lambda(t, z, d_s)$. Within each family of covariate-adjusted angular densities, likelihood ratio tests (LRT) are performed to select the most adequate VGAM for the dependence parameters. Table 2 shows the best models in each of the three families of angular densities. All the considered covariates have a significant effect on the strength of dependence between extreme high temperatures in Montana and Zermatt. For the covariate-dependent Dirichlet model, the covariates affect the dependence parameters α and β differently. However, these parameters lack interpretability, and Coles and Tawn [1994] mention the quantities $(\alpha + \beta)/2$ and $(\alpha - \beta)/2$ that can be interpreted as the strength and asymmetry of the extremal dependence, respectively. In this case, the best Dirichlet dependence model found in Table 2 is such that both the intensity and the asymmetry of the dependence are affected by time, NAO, and day in season.

TABLE 2 Selected models in each family of angular densities along with their AICs. The link functions g are the logit function for the logistic model and the logarithm function for the Dirichlet and the Hüsler–Reiss models. The functions $\hat{f}$ with subscripts t, z, and d_s are fitted smooth functions of time, NAO, and day in season, respectively

Covariate-adjusted angular density	VGAM	AIC
Logistic	$\hat{\alpha}(t, z, d_s) = g^{-1}\{\hat{\alpha}_0 + \hat{f}_t(t) + \hat{f}_z(z) + \hat{f}_{d_s}(d_s)\}$	−280.14
Dirichlet	$\hat{\alpha}(z) = g^{-1}\{\hat{\alpha}_0 + \hat{f}_z(z)\}$ $\hat{\beta}(t, d_s) = g^{-1}\{\hat{\beta}_0 + \hat{f}_t(t) + \hat{f}_{d_s}(d_s)\}$	−290.04
Hüsler–Reiss	$\hat{\lambda}(t, z, d_s) = g^{-1}\{\hat{\lambda}_0 + \hat{f}_t(t) + \hat{f}_z(z) + \hat{f}_{d_s}(d_s)\}$	−274.52

The best models in the studied angular density families are then compared by means of the AIC (see Section 3.2) displayed in Table 2. The Dirichlet model with $\alpha(z)$ and $\beta(t, d_s)$ parameters has the lowest AIC and is hence selected. This suggests the presence of asymmetry in the dependence of extreme high temperatures between Montana and Zermatt. Figure 9 shows the fitted smooth effects of the covariates on the extremal coefficient—constructed via the covariate-adjusted extremal coefficient as in (10)—that lies between 1 for perfect extremal dependence and 2 for perfect extremal independence.

A decrease in the extremal coefficient, or equivalently an increase in the extremal dependence between high winter temperatures in Montana and Zermatt, is observed from 1988 until 2006. This change might be explained first by a warm phase of very pronounced and persistent warm anomalies during the winter season, which occured countrywide from 1988 to 1999 [Jungo and Beniston, 2001], and then by an exceptionally warm 2006/2007 winter that took place

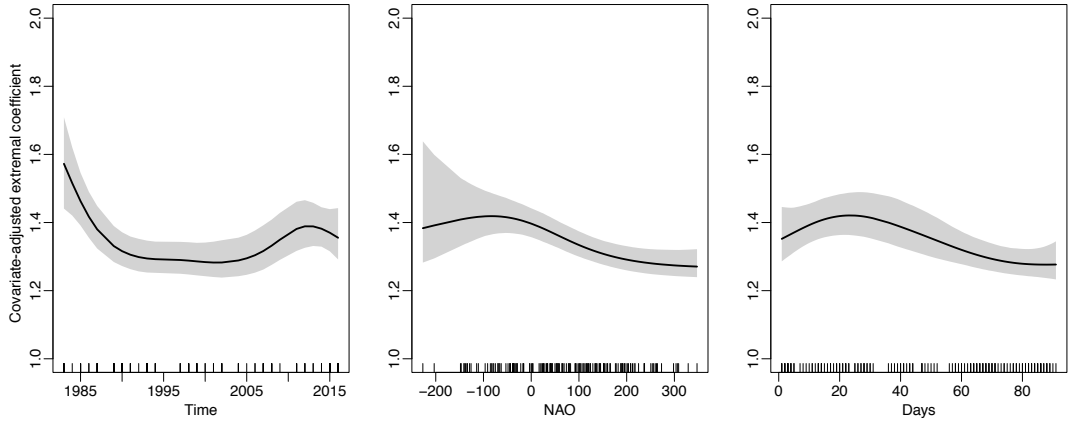

FIGURE 9 Fitted smooth effects for the extremal coefficient under the Dirichlet model of Table 2 along with their associated 95% (pointwise) asymptotic confidence bands.

in Europe Luterbacher et al. [2007]. Regarding the NAO effect, as expected, we observe an increase in the extremal dependence during the positive phase of NAO that has a geographically global influence on the Alps and results in warmer and milder winters, as depicted by Beniston [1997]. In terms of the very negative NAO values (less than −100), there is an important uncertainty due to the corresponding small amount of joint extreme high temperatures (8%). The right panel of Figure 9 suggests an increase in the extremal dependence around mid-December. This evidence also seems compatible with the countrywide findings by Beniston [1997], who claims that

> *"The anomalously warm winters have resulted from the presence of very persistent high pressure episodes which have occurred essentially during periods from late Fall to early Spring."*

Threshold selection entails a bias–variance tradeoff, and a standard practice in statistics of extremes is to assess the stability of the inference over a range of neighbouring quantiles at which the threshold is set. Here, the sensitivity of the dependence modelling to the choice of the radial threshold (95% quantile) is assessed by fitting the Dirichlet model of Table 2 to the angular observations corresponding to a radial component exceeding its 90%, 93%, and 97% quantile. The fitted smooth effects of the covariates on the extremal coefficient are reported in the Supplementary Materials and are found to be essentially unaffected by the choice of the radial threshold.

5.2.2 Dependence of extreme low winter temperatures

The effects of the covariates time, NAO, and day in season on the dependence between extreme cold winters in Montana and Zermatt are now tested by fitting the bivariate angular densities of Section 2.2. Within each of the logistic, Dirichlet, and Hüsler–Reiss families, LRTs are performed, and the selected models are displayed in Table 3. The explanatory covariates have different effects on the extremal dependence, depending on the family of angular densities. The AICs for the fitted models are quite close, and the asymmetric Dirichlet model has the lowest AIC and is hence the retained model. As opposed to the extremal dependence between warm winters in the two mountain sites, the NAO has a non-significant effect on the extremal dependence between cold winters. This might be explained

TABLE 3 Selected models in each family of angular densities along with their AICs. The link functions g are the logit function for the logistic model and the logarithm function for the Dirichlet and the Hüsler–Reiss models. The functions $\hat{f}$ with subscripts t and d_s are fitted smooth functions of time and day in season, respectively

Covariate-adjusted angular density	VGAM	AIC
Logistic	$\hat{\alpha} \equiv g^{-1}(\hat{\alpha}_0)$	−358.49
Dirichlet	$\hat{\alpha}(t, d_s) = g^{-1}\{\hat{\alpha}_0 + \hat{f}_t(t) + \hat{f}_{d_s}(d_s)\}$ $\hat{\beta}(d_s) = g^{-1}\{\hat{\beta}_0 + \hat{f}_{d_s}(d_s)\}$	−363.49
Hüsler–Reiss	$\hat{\lambda}(t) = g^{-1}\{\hat{\lambda}_0 + \hat{f}_t(t)\}$	−362.05

by the fact that high values of the NAO index will affect the frequency of extreme low winter temperatures (less extremes) and hence the marginal behaviour of the extremes at both sites, but not necessarily the dependence of the extremes between these sites [Beniston, 2004, sec. 7.3.2].

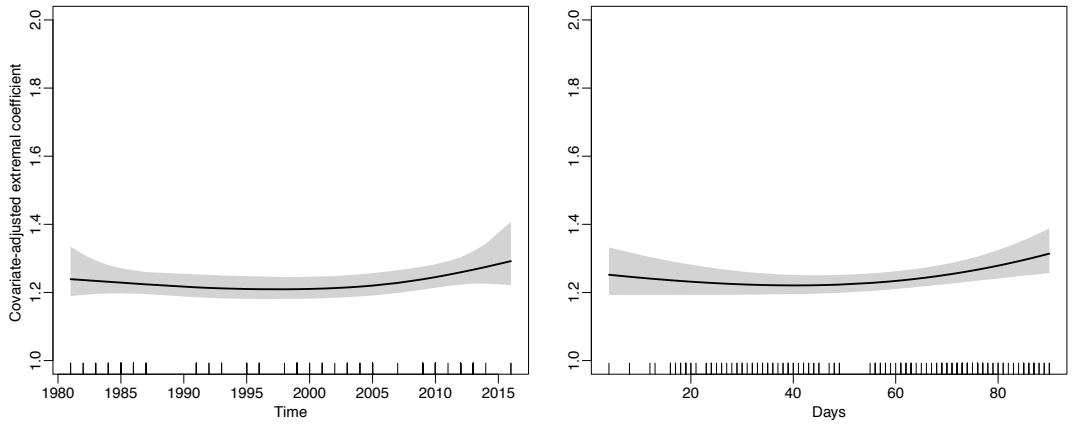

FIGURE 10 Fitted smooth effects for the extremal coefficient under the Dirichlet model of Table 3 along with their associated 95% (pointwise) asymptotic confidence bands.

Figure 10 shows the fitted smooth effects of time and day in season. A sensitivity analysis of these effects to the radial threshold choice is conducted in Figure 1 of the Supplementary Materials and shows that the effects of the considered covariates on the extremal dependence are unaltered by changing the radial threshold; thus we only present here the results yield from setting the radial threshold at the 95% quantile. The extremal dependence between low winter temperatures in Montana and Zermatt is high, regardless of the values taken by the covariates t and d_s. The range of values of the extremal coefficient observed in Figure 10 is in line with the findings of Davison et al. [2013], where the value of the extremal coefficient for the dependence between extreme low winter temperatures (in Switzerland) is around 1.3 for pairs of resorts separated by up to 100km. Overall, the extremal coefficient is lower in the extreme low winter temperatures than in the extreme high winter temperatures. This could be explained by the fact that minimum winter temperatures are usually observed overnight when the atmosphere is purer and not affected by local sunshine effects and hence is more favourable to the propagation over space of cold winter spells.

A decrease in the extremal dependence is observed from around 2007 and results in values of the extremal coefficient that are comparable to those obtained under the warm winter spells scenario (see Figure 9). This can be explained by a decrease in the intensity of the joint extreme low temperatures, that is, milder joint extreme low temperatures, occurring during the last years of the analysis, as can be observed in Figure 11. The right panel of Figure 10 highlights a decrease in the extremal dependence when approaching spring. This effect can be explained by the fact that mountains often produce their own local winds, as can be seen for instance from

`www.morznet.com/morzine/climate/local-climate-in-the-alps`

These warm dry winds are mostly noticeable in spring and are called Foehn in the Alps. Local effects obviously lead to a decrease of extremal dependence between the two resorts.

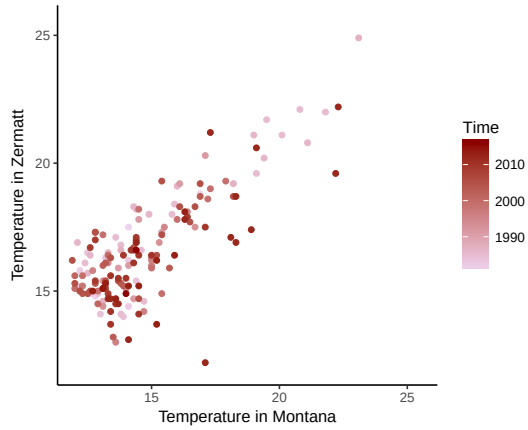

FIGURE 11 Scatterplot of (minus) extreme low winter temperatures (in °C) in Montana and Zermatt.

6 | FINAL REMARKS

In this paper, we have introduced a sturdy and general approach to model the influence of covariates on the extremal dependence structure. Keeping in mind that extreme values are scarce, our methodology borrows strength from a parametric assumption and benefits directly from the flexibility of VGAMs. Our non-linear approach for covariate-varying extremal dependences can be regarded as a model for conditional extreme value copulas—or equivalently as a model for nonstationary multivariate extremes. An important advantage over existing methods is that our model profits from the VGAM framework, allowing the incorporation of a large number of covariates of different types (continuous, factor, etc) as well as the possibility for the smooth functions to accommodate different shapes. The fitting procedure is an iterative ridge regression, the implementation of which is based on an ordinary N–R type algorithm that is available in many statistical software. An illustration is provided in the R code that can be downloaded from `https://github.com/lindamhalla/Regression_type_models_for_extremal_dependence`.

The method paves the way for novel applications, as it is naturally tailored for assessing how covariates affect dependence between extreme values—and thus it offers a natural approach for modelling conditional risk. Conceptually, the proposed approach is valid in high dimensions. Yet, as for the classical setting without covariates, the number of parameters would increase quickly with the dimension and additional complications would arise. Relying on composite likelihoods [Padoan et al., 2010] instead of the full likelihood seems to represent a promising path for future extensions of the proposed methodology in a high-dimensional context.

Theoretical developments on threshold selection for nonstationary multivariate extremes are a natural avenue for future research. Threshold selection on the (stationary) multivariate setting is still an area of active research; recently, [Wan and Davis, 2018, in press] proposed an approach for the multivariate setting by resorting to the well-known fact that the radial and angular components are independent in the limit. Formal threshold selection for nonstationary multivariate extremes entails the challenge that the dependence structure changes according to a set of covariates, and thus requires the need for focusing on models based on angular surfaces rather than angular densities.

Acknowledgement

We thank the Editor and the Reviewers for insightful comments on a previous version of the paper. M. de Carvalho acknowledges funding from the FCT (Fundação para a Ciência e a Tecnologia, Portugal) through the projects UID/-MAT/00006/2013 and PTDC/MAT-STA/28649/2017. We thank participants of Workshop 2017, EPFL, for discussions and comments. We also would like to thank Paul Embrechts for his constant encouragement.

Supporting information

Additional information for this article is available online including the proof of Theorem 1 and supplementary empirical analyses.

References

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D. and Ferro, C. (2004) *Statistics of extremes: Theory and applications*. New York: Wiley.

Beniston, M. (1997) Variations of snow depth and duration in the Swiss Alps over the last 50 years: Links to changes in large-scale climatic forcings. *Climatic Change*, **36**, 281–300.

— (2004) *Climatic change and its impacts: an overview focusing on Switzerland*. Advances in Global Change Research. Netherlands: Springer.

— (2005) Warm winter spells in the Swiss Alps: Strong heat waves in a cold season? A study focusing on climate observations at the Saentis high mountain site. *Geophysical Research Letters*, **32**, 1–5.

— (2007) Linking extreme climate events and economic impacts: Examples from the Swiss Alps. *Energy Policy*, **35**, 5384–5392.

Beniston, M. and Rebetez, M. (1996) Regional behavior of minimum temperatures in Switzerland for the period 1979–1993. *Theoretical and Applied Climatology*, **53**, 231–243.

Bienvenüe, A. and Robert, C. Y. (2017) Likelihood inference for multivariate extreme value distributions whose spectral vectors have known conditional distributions. *Scandinavian Journal of Statistics*, **44**, 130–149.

Boldi, M.-O. and Davison, A. C. (2007) A mixture model for multivariate extremes. *Journal of the Royal Statistical Society, Series B*, **69**, 217–229.

Castro, D. and de Carvalho, M. (2017) Spectral density regression for bivariate extremes. *Stochastic Environmental Research and Risk Assessment*, **31**, 1603–1613.

Castro, D., de Carvalho, M. and Wadsworth, J. (2018) Time-varying extreme value dependence with application to leading european stock markets. *Annals of Applied Statistics*, **12**, 283–309.

Chavez-Demoulin, V. and Davison, A. C. (2005) Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society, Series C*, **54**, 207–222.

Coles, S. (2001) *An introduction to statistical modeling of extreme values*. London: Springer.

Coles, S. and Tawn, J. A. (1991) Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B*, **53**, 377–392.

— (1994) Statistical methods for multivariate to structural design extremes: An application to structural design. *Journal of the Royal Statistical Society, Series C*, **43**, 1–48.

Cooley, D., Davis, R. A. and Naveau, P. (2010) The pairwise Beta distribution: A flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis*, **101**, 2103–2117.

Davison, A. C. (2003) *Statistical models*. Cambridge, UK: Cambridge University Press.

Davison, A. C. and Gholamrezaee, M. M. (2011) Geostatistics of extremes. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **468**, 581–608.

Davison, A. C., Huser, R. and Thibaud, E. (2013) Geostatistics of dependent and asymptotically independent extremes. *Mathematical Geosciences*, **45**, 511–529.

de Carvalho, M. (2016) Statistics of extremes: Challenges and opportunities. In *Extreme events in finance: A handbook of extreme value theory and its applications* (ed. F. Longin). Hoboken: Wiley.

de Carvalho, M. and Davison, A. C. (2014) Spectral density ratio models for multivariate extremes. *Journal of the American Statistical Association*, **109**, 764–776.

de Carvalho, M., Oumow, B., Segers, J. and Warchoł, M. (2013) A Euclidean likelihood estimator for bivariate tail dependence. *Communications in Statistics—Theory and Methods*, **42**, 1176–1192.

Dombry, C., Éyi Minko, F. and Ribatet, M. (2013) Conditional simulation of max-stable processes. *Biometrika*, **100**, 111–124.

Drees, H. and Kaufmann, E. (1998) Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, **75**, 149–172.

Eastoe, E. F. and Tawn, J. A. (2009) Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society, Series C*, **58**, 25–45.

Einmahl, J. H. J., de Haan, L. and Li, D. (2006) Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *The Annals of Statistics*, **34**, 1987–2014.

Einmahl, J. H. J., Li, J. and Liu, R. Y. (2009) Thresholding events of extreme in simultaneous monitoring of multiple risks. *Journal of the American Statistical Association*, **104**, 982–992.

Escobar-Bach, M., Goegebeur, Y. and Guillou, A. (2016) Local robust estimation of the Pickands dependence function. Hal-01340166.

Fisher, R. A. and Tippett, L. H. C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, 180–190. Cambridge University Press.

Greenstadt, J. (1967) On the relative efficiencies of gradient methods. *Mathematics of Computation*, **21**, 360–367.

Gudendorf, G. and Segers, J. (2012) Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference*, **142**, 3073–3085.

de Haan, L. and Resnick, S. I. (1977) Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **40**, 317–337.

Hanson, T. E., de Carvalho, M. and Chen, Y. (2017) Bernstein polynomial angular densities of multivariate extreme value distributions. *Statistics and Probability Letters*, **128**, 60–66.

Huang, X. (1992) *Statistics of Bivariate Extreme Values*. Ph.D. thesis, Tinbergen Institute Research Series.

Huser, R., Davison, A. C. and Genton, M. G. (2016) Likelihood estimators for multivariate extremes. *Extremes*, **19**, 79–103.

Hüsler, J. and Reiss, R.-D. (1989) Maxima of normal random vectors: Between independence and complete dependence. *Statistics and Probability Letters*, **7**, 283–286.

Jungo, P. and Beniston, M. (2001) Changes in the anomalies of extreme temperature anomalies in the 20th century at Swiss climatological stations located at different latitudes and altitudes. *Theoretical and Applied Climatology*, **69**, 1–12.

Knight, K. (2000) *Mathematical Statistics*. Boca Raton: Chapman & Hall/CRC Press.

Kotz, S. and Nadarajah, S. (2000) *Extreme value distributions: Theory and applications*. London: Imperial College Press.

Luterbacher, J., Liniger, M. A., Menzel, A., Estrella, N., Della-Marta, P. M., Pfister, C., Rutishauser, T. and Xoplaki, E. (2007) Exceptional European warmth of autumn 2006 and winter 2007: Historical context, the underlying dynamics, and its phenological impacts. *Geophysical Research Letters*, **34**, 1–6.

Marcon, G., Padoan, S. A. and Antoniano-Villalobos, I. (2016) Bayesian inference for the extremal dependence. *Electronic Journal of Statistics*, **10**, 3310–3337.

Marra, G. and Wood, S. (2011) Practical variable selection for generalized additive models. *Computational Statistics Data Analysis*, **55**, 2372 – 2387.

Marra, G. and Wood, S. N. (2012) Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, **39**, 53–74.

Mhalla, L., Chavez-Demoulin, V. and Naveau, P. (2017) Non-linear models for extremal dependence. *Journal of Multivariate Analysis*, **159**, 49–66.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, **105**, 263–277.

Pauli, F. and Coles, S. (2001) Penalized likelihood inference in extreme value analyses. *Journal of Applied Statistics*, **28**, 547–560.

Pickands, J. (1981) Multivariate extreme value distributions. In *Proc. 43rd Session of the International Statistical Institute*, 859–878.

R Development Core Team (2016) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Resnick, S. I. (1987) *Extreme values, regular variation and point processes*. New York: Springer.

Ruppert, D. (2002) Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, **11**, 735–757.

Sabourin, A. and Naveau, P. (2014) Bayesian Dirichlet mixture model for multivariate extremes: A re-parametrization. *Computational Statistics and Data Analysis*, **71**, 542–567.

Tawn, J. A. (1990) Modelling multivariate extreme value distributions. *Biometrika*, **77**, 245–253.

Vatter, T. and Chavez-Demoulin, V. (2015) Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, **141**, 147–167.

Wadsworth, J. L. and Tawn, J. A. (2013) A new representation for multivariate tail probabilities. *Bernoulli*, **19**, 2689–2714.

Wan, P. and Davis, R. A. (2018, in press) Threshold selection for multivariate heavy-tailed data. *Extremes*, 1–36.

Wang, H. and Tsai, C.-L. (2009) Tail index regression. *Journal of the American Statistical Association*, **104**, 1233–1240.

Wood, S. (2017) *Generalized additive models: An introduction with R*. Chapman and Hall/CRC, 2 edn.

Yee, T. W. (2015) *Vector generalized linear and additive models: With an implementation in R*. New York: Springer, 1st edn.

— (2016) Comment on: "Smoothing parameter and model selection for general smooth models," by Wood, S. N., Pya, N., and Säfken, N. *Journal of the American Statistical Association*, **111**, 1565–1568.

— (2017) *VGAM: Vector Generalized Linear and Additive Models*. URL: `https://CRAN.R-project.org/package=VGAM`. R package version 1.0-4.

Yee, T. W. and Stephenson, A. G. (2007) Vector generalized linear and additive extreme value models. *Extremes*, **10**, 1–19.

Yee, T. W. and Wild, C. J. (1996) Vector generalized additive models. *Journal of the Royal Statistical Society, Series B*, **58**, 481–493.